

The international journal of science / 5 March 2020

nature

SATURATION POINT

The ability of tropical forests to sequester CO₂ is in decline

Core issue

The quest to discover what lies at the heart of neutron stars

Machine vision

Image-sensor array is its own neural network

Walking tall

The transverse arch that helped put humans on their feet

Vol. 579, No. 7797
nature.com

The pandemic question

For now, the World Health Organization is right to emphasize containment of the new coronavirus.

More than 3,000 recorded deaths and 90,000 confirmed infections, and the numbers are still rising. The coronavirus that causes COVID-19 has spread to more than 70 countries, with more nations being affected daily. As new clusters emerge, all eyes are on the World Health Organization (WHO).

Last week, the agency held back from describing the outbreak as a pandemic – usually understood to mean the spread across multiple regions of a disease that cannot be contained (see p. 12). The WHO's decision was based partly on the fact that most of the virus's global spread can still be traced to countries that have experienced large outbreaks, such as China, Iran, Italy and South Korea. There are signs – in China, for example, where the spread of disease seems to be slowing – that the virus could yet be contained if the right measures are put in place.

Another argument for not using 'pandemic' is that much of the world is already on maximum alert. Countries are restricting travel; borders are being sealed; schools and public buildings are being shuttered; and gatherings, including research conferences, are being called off (see p. 13). Moreover, a huge effort is being made to trace and track new outbreaks; researchers are collaborating across borders to determine and share virus genome sequences; vaccine development is under way; and many journals are making all related research and data open access.

David Heymann, an infectious-disease epidemiologist at the London School of Hygiene and Tropical Medicine who led the WHO's response to severe acute respiratory syndrome (SARS) in 2003, told *Nature* that he is not advising the WHO to call it a pandemic at this point – partly because the virus is not spreading in the same way as the pandemics of the twentieth century, which claimed millions of lives. There are also the economic implications to consider. Even without the virus being described as a pandemic, the values of stocks and shares have fallen sharply and some economies are at risk of recession.

But the virus is still spreading daily, and more previously undetected clusters will probably be found, such as those recently discovered in the United States. Marc Lipsitch, an infectious-disease epidemiologist at the Harvard T.H. Chan School of Public Health in Boston, Massachusetts, told *Nature* that "under almost any reasonable definition of pandemic, there's now evidence of it happening".

Part of the difficulty for the WHO is that the impact of a pandemic declaration in previous disease outbreaks is


Much of the world is already on maximum alert."

hard to assess, because there are few examples to go on. The 2002–03 outbreak of the SARS coronavirus, which killed 774 people (out of a total of 8,098 infections, spread across some two dozen countries), was not described as a pandemic by the WHO. Neither was the 2014–16 Ebola outbreak, which affected three countries in West Africa, and resulted in 28,616 infections and 11,310 deaths.

In the case of SARS, Heymann says, most transmission occurred in clusters of infected health-care workers and hospital patients, and in the families of health-care workers, with occasional transmission in the wider community. A similar pattern was seen in the early outbreaks of the new coronavirus in China, and is now occurring in other countries. SARS "was not a pandemic in the sense of pandemic influenza or cholera, where transmission was more generalized", Heymann says.

The WHO did declare the 2009–10 H1N1 influenza outbreak a pandemic, partly to trigger the release of funding for vaccine production. At present, however, there is no vaccine against the virus that causes COVID-19. The agency has also stopped using the definition of pandemic that it used at that time. On that occasion, some people criticized the agency for over-reacting – initial estimates of deaths were about 18,600. But that number looks to have been an undercount, and revised estimates of fatalities in the first year that the virus circulated range from 150,000 to 575,000. There were 61 million infections in the United States alone (L. Simonsen *et al.* *PLoS Med.* **10**, e1001558; 2013).

On previous occasions, much of the WHO's work involved persuading reluctant governments to acknowledge the severity of an infectious-disease outbreak. Fortunately, that has changed with the virus that causes COVID-19.

If past outbreaks are a guide, we are only in the foothills of a new disease that could continue to spread for many more months. All countries must put in place containment measures. But the p-word should remain on the table. If the virus spread accelerates, it may be necessary to use it.

A warning from the forests of Africa and the Amazon

Carbon analysis suggests faster emissions reductions are needed.

Astropical forests grow, they pull carbon dioxide out of the atmosphere – one of their many services to humanity and the planet.

Decades of measurements in hundreds of plots in Africa and South America show how tropical trees such as Brazil nut (*Bertholletia excelsa*) and kapok (*Ceiba pentandra*) absorbed as much as 4.4 billion

tonnes of carbon dioxide annually in the 1990s and early 2000s. That's enough to more than offset the European Union's carbon emissions during the same period.

This effect is baked into many of the climate models that researchers use to project future global-warming scenarios. However, a study published in *Nature* this week suggests that the benefits from this tropical carbon 'sink' might be fleeting (W. Hubau *et al. Nature* **579**, 80–87; 2020). And that could mean the international community will need to pledge yet faster emissions reductions if the world is to limit global warming to below 2 °C, in line with the 2015 Paris climate agreement.

An international team led by geographers from the University of Leeds, UK, reports on page 80 that the Amazon rainforest has been absorbing less atmospheric carbon each year since the early 1990s. Forests in Africa have also been absorbing less atmospheric carbon since around 2015. This is due in large part to rising tree mortality.

Trees are dying, the researchers found, because temperatures are rising and drought is increasing, a trend that is likely to continue as greenhouse gases build up. A decade from now, Africa's carbon sink will be 14% lower compared with 2010–15. The Amazonian carbon sink is on course to disappear completely by 2035. If that happens it will result in more carbon dioxide in the atmosphere, and therefore more global warming.

As we reported in a Feature last week, the Amazon's 5 million square kilometres look more precarious than ever (see *Nature* **578**, 505–507; 2020). Average temperatures in this rainforest, which spans nine countries, have risen by 1–1.5 °C over the past century; there have been three severe droughts since 2005 and tree clearing has shrunk the forest by 15% since the 1970s. Brazil, once praised for its efforts in slowing deforestation, lost 10,000 square kilometres last year – the largest drop for a decade. A ten-year ban on planting sugar cane in the Amazon was lifted last November; and a bill to regulate oil and mining exploration is making its way to the national congress, Brazil's parliament.

In September, independent researchers from the region formed a science panel to propose what needs to be done to conserve the Amazon. The panel hasn't yet completed its report, but its overarching message cannot be in doubt: Brazil and other tropical nations need to halt deforestation and promote new forests in degraded – and often abandoned – lands.

At November's summit of the United Nations Framework Convention on Climate Change in Glasgow, UK, participating countries will be expected to redouble their pledges to meet the Paris climate agreement's goals. If tropical carbon sinks can no longer be fully relied upon to help reach that target, it means more ambitious decarbonization will be needed.

At the same time, the lesson for governments around the world is clear enough: tropical forests are working for humanity – and for countless other creatures. To protect them, humanity must halt both deforestation and global warming.

“
The Amazon
carbon sink
is on course
to disappear
completely
by 2035.”

China changes tack

A new researcher-evaluation system must not reduce international collaborations.

China's researchers and research institutions are evaluated, ranked and funded according to their record of publishing in journals covered by the Science Citation Index (SCI), an international database of articles and citation records for around 9,000 journals.

The number of articles in these journals by authors at Chinese institutions increased nearly fourfold between 2009 and 2019. In that time, China's researchers have increased international collaborations, which have helped them secure international publications. But there have been concerns that widespread use of publication metrics incentivizes lower-quality work, as does the fact that some institutions pay bonuses to those publishing in journals.

But that might be about to change. Last month, the Chinese government ordered institutions to stop promoting or recruiting solely on the basis of number of papers or citations, and to end publishing bonuses (see page 18).

Research will still be evaluated, but institutions have until the end of July to propose new indicators. An alternative system will need to measure research quality and innovation, and whether something represents a significant advance or helps to solve an important societal problem. Evaluators will need to rely more on peer judgement, and, crucially, researchers must consider publishing in non-SCI-indexed journals.

The change is significant, and intended to meet two important government objectives. First, it is designed to help root out plagiarism, self-citation and colleagues citing each others' work to boost their citations. Second, it is aimed at boosting China's own research-publishing industry, which the government has wanted to do – but which is difficult if the best research is published internationally.

To enable more domestic research publishing, the government last year allocated one billion yuan (US\$143 million) over 5 years to improve the standards of some 280 Chinese journals, most of which publish in English. These journals have been ranked, with each of the top 22 receiving between one million and five million yuan annually to help them attract a higher standard of submissions, not only from China, but from around the world.

When this policy was announced, it wasn't known how the publishers would use their subsidy or how the government would measure success. The answers to both questions are now clearer.

China's government is urging its researchers to play their part by publishing in home-grown journals. That is important, not least because it will make science more accessible in China. But in setting up the new evaluation system, the government must be careful to protect the collaborations – and the relationships – that came with the old.

World view

Extended US travel ban harms global science

From preparing for pandemics to boosting crop yields, Nigerian scientists who work and train abroad are making the world safer. Now that's under threat.

In the first weeks of the COVID-19 outbreak, I worked with others to set up capabilities to detect the disease at our National Reference Laboratory in Abuja, and at the Lagos University Teaching Hospital. These cities are the main points of entry into Nigeria – the most populous country in Africa. We detected the country's first case of the coronavirus on 27 February, in Lagos. That Nigerian researchers could act without having to wait for external partners has made my country, and the world, safer.

But I fear that our capacity to act in this way is under threat. Much of the expertise that my colleagues and I drew on came from our experiences training and working abroad. I've spent time at Graz in Austria; Yale University in New Haven, Connecticut; and the Harvard Stem Cell Institute in Cambridge, Massachusetts. Chikwe Ihekweazu, who leads the Nigeria Centre for Disease Control and whose support was crucial to setting up the country's detection facilities, also trained abroad (in his case, in Germany, where an anti-immigrant attack took place last month).

This kind of expertise goes beyond global health: a Nigerian geneticist who trained at Cornell University in Ithaca, New York, is using genotyping to selectively breed our indigenous cows so that they can resist disease and produce high-quality milk. A Nigeria-based project to improve cassava could save the crop in this country, and in the rest of the world. Nigerian software engineers are developing medical-records databases and e-commerce applications that will strengthen my country's health and economy. In short, many scientists from Nigeria who trained in the United States, the United Kingdom and elsewhere are returning home, inspired to improve people's lives. We should encourage more of this.

Yet our ability to travel is contracting. On 21 February, people from Nigeria, and also from Myanmar, Eritrea, Kyrgyzstan, Sudan and Tanzania, joined those from Libya, Iran, Somalia, Syria and Yemen – as well as from North Korea and Venezuela – in facing restricted entry to the United States. The extended travel ban does not officially affect students or highly specialized workers. Practically, however, it means that scientists will find immigration more difficult and that, if they do get visas, their close family members will not be able to join them.

When I learnt the news, I was stunned. In 2018, according to the New American Economy research group, more than 645,000 immigrants arrived in the United States from the 6 countries recently added to the travel ban; more than half

The quality of training we received abroad gave us the confidence and expertise to work with leading scientists on COVID-19."

Nnaemeka Ndodo is chief molecular bioengineer at the National Reference Laboratory for Nigeria's Centre for Disease Control in Abuja. nnaemeka.ndodo@ncdc.gov.ng



By Nnaemeka Ndodo

of these immigrants were Nigerian. I do not see how my country can be considered a threat to the United States. The two countries have close diplomatic, economic and social ties. The United Nations estimated that some 1.24 million people made up the Nigerian diaspora in 2017, with the United States attracting the vast majority of emigrants, and numbers were growing.

These people have boosted prosperity in the countries that received them. Nigerians form one of the most educated immigrant communities in the United States and paid more than US\$4 billion in federal, state and local taxes in 2018. They are an asset, not a liability. Whatever the rationale for the baffling expansions to the ban, the consensus is that it will disconnect Nigerians in the United States from their families and make for a more hostile environment.

Even without such moves, travelling can be difficult. Just the sight of my country's green passport can bring excessive scrutiny. On my very first international trip, airport officials in a major European city searched my luggage. When they found an anatomy book, they decided I must be concealing drugs in my body. So they forced me to sign a document at gunpoint, took me from the airport without a coat (I had never experienced such cold) and performed X-rays (another first) that found no contraband. When I demanded an apology, they said I should blame my maltreatment on compatriots selling narcotics. I was ready to return to Nigeria immediately to avoid future ill treatment. I am glad now that a colleague convinced me otherwise, and I continued my journey to train in molecular bioengineering.

Already, many young Nigerian scientists wishing to study in the United States are repeatedly denied the sorts of training opportunity that I had, their visas being turned down for one reason after another. This is so short-sighted: if a new human or crop pathogen arises here, we need top molecular biologists to detect it. For that, Nigerian scientists need to be in contact with their colleagues in other countries. The quality of training we received abroad gave us the confidence and expertise to work with leading scientists on COVID-19, to understand its genetic-sequence data and to use World Health Organization-approved target sequences (DNA primers and probes for amplifying any viral material) to set up early-detection capacity in our country.

I am not saying that the ability to travel freely will on its own result in world-class institutions in Nigeria. My country's government needs to do more to build science here, for instance, by providing more (and more reliable) funding, and establishing a dependable power grid.

The visa ban, no matter its justification, will worsen an already bad situation. It can only have a damaging effect on the Nigerian economy and on Nigerian scientists. And that – in an increasingly connected world – will hurt all of us.

News in brief

CORONAVIRUS ENTERS DANGEROUS NEW PHASE

The new coronavirus has spread to more than 70 nations and the total number of infections worldwide had passed 90,000 as *Nature* went to press (see 'Rapid spread').

Researchers have warned that the surge in outbreaks outside China, where the virus emerged and most cases have occurred, means that the coronavirus is becoming unstoppable.

The World Health Organization has resisted describing the situation as a pandemic. Director-general Tedros Adhanom Ghebreyesus said on 2 March that there was still a chance of containing the virus. Mike Ryan, director of the WHO's emergencies programme, said that using the word pandemic would mean that efforts to contain and slow the spread of the virus have failed, which has proved untrue in China, Singapore and other regions.

But other scientists say the surge in international cases marks a tipping point. "I think the epidemiological conditions for a pandemic are met," says Marc Lipsitch,

an infectious-disease epidemiologist at the Harvard T.H. Chan School of Public Health in Boston, Massachusetts.

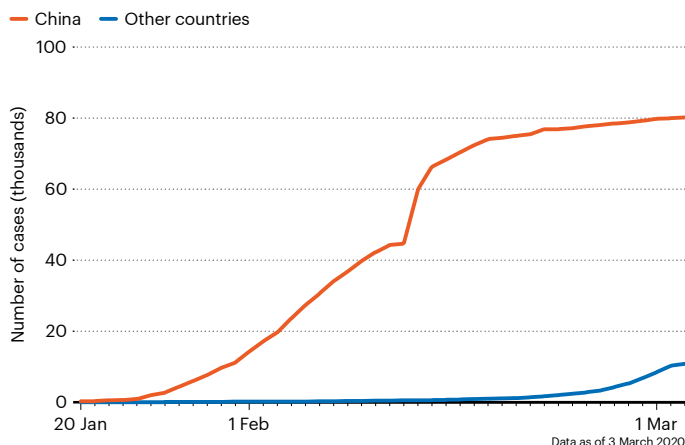
He and others say that although containment measures seem to have kept outbreaks from escalating outside China for more than a month, such procedures might soon become unfeasible on a broader scale. Those efforts have involved quickly identifying infected people and their close contacts, and isolating them to prevent further transmission.

"We've got to think more carefully about what measures might be sustainable in terms of reducing transmission without shutting down cities completely and stopping people from moving," says Ben Cowling, an infectious-disease epidemiologist at the University of Hong Kong.

The efforts include 'social distancing', which reduces the average chances that uninfected people will encounter an infected person. But some epidemiologists say too little is known about the outbreak to deploy this effectively.

RAPID SPREAD

The new coronavirus has infected more than 90,000 people globally and spread to more than 70 countries. The vast majority of cases — some 80,000 — are in China, where the pathogen emerged.



BETELGEUSE STARTS TO BRIGHTEN AGAIN

After a mysterious four-month fading streak, the star known as Betelgeuse could be on its way to regaining its shine.

Easily recognizable as the right 'shoulder' in the constellation Orion, Betelgeuse is normally one of the ten brightest stars in the night sky. But it began getting dimmer in October last year, and by mid-February it had lost more than two-thirds of its brilliance — a difference noticeable to the naked eye. The star now appears to be recovering, and has brightened by around 10% from its dimmest point.

Astronomers have proposed several explanations for the dimming. One is the emergence of a large, unusually cool convection cell — a blob of cooling plasma on its surface. Another is that the star could be moving behind a dust cloud.

Some have speculated that the star's erratic swings in brightness mean it might be approaching the end of its life. Astrophysicists predict Betelgeuse will end in a supernova sometime in the next 100,000 years. But what happens right before a star explodes in this way is unknown, and the exact timing of the fiery end is impossible to predict.



Next stop, the twilight zone



It is home to a majority of the marine fish biomass and helps to remove an estimated 4 billion tonnes of carbon dioxide from the atmosphere each year. Now, scientists are gearing up to dive into the twilight zone, the largely unexplored ocean layer 200–1,000 metres deep that some worry is threatened by a changing climate and increased pressure from fishing.

As part of a US\$25-million mission, NASA will travel to the North Atlantic in April to study the movement of carbon between the atmosphere and the deep ocean. Others will join the expedition thanks to a collaborative venture unveiled at the American Geophysical Union's ocean-science meeting in San Diego, California, last week.

"This is literally the biggest investment ever made in the twilight zone," says Dave Siegel, an oceanographer at the University of California, Santa Barbara. He is heading the NASA mission, dubbed Export Processes in the Ocean from Remote Sensing, or EXPORTS. The addition of a network of collaborators promises to bolster data-sharing and coordination with other research efforts around the world. "If we can federate, we can help each other," he says.

CORONAVIRUS NIXES MASSIVE PHYSICS MEETING

One of the world's biggest scientific conferences – the March Meeting of the American Physical Society (APS) – was cancelled just before it was scheduled to begin in Denver, Colorado, for fear of contributing to the spread of coronavirus.

Hundreds of registered participants had already arrived in Denver when they received an e-mail from the APS on 29 February. The week-long meeting was set to begin on 2 March, with more than 11,000 attendees. APS leaders said that a major factor was the decision by the US Centers for Disease Control and Prevention to issue the highest level of travel warning for Italy and South Korea, where coronavirus outbreaks are growing rapidly. The warning includes a recommendation to avoid all non-essential trips.

Still, physicists are finding ways to get the word out about their research, despite the cancellation. Some will record their talks and upload them to virtualmarchmeeting.com, a website quickly set up for this purpose. The APS itself says it will provide a platform for sharing presentations, and is asking registrants to submit links to their talks. Some scientists in Denver are holding informal get-togethers for their disciplines, a practice called unconferencing.

"It was clear that nothing formal was possible, like recreating the whole meeting virtually", so speakers were invited to post their own links to an online spreadsheet instead, says Karen Daniels, a physicist at North Carolina State University in Raleigh who is leading one disciplinary effort.



WHY RATS ARE NEW YORKERS TOO

An analysis of the genomes of New York City rats is offering clues to the rodents' ability to thrive in urban jungles. Researchers identified dozens of areas of the rat genome that were specific to animals in New York – including several linked to diet, behaviour and mobility.

"I can't help but be amazed by the ways that rats have adapted to urban environments," says Arbel Harpak, a population geneticist at the city's Columbia University, who co-led the study (A. Harpak *et al.* Preprint at [bioRxiv http://doi.org/dnxj; 2020](https://doi.org/10.1101/2020.03.10.331111)).

Harpak's team sequenced the full genomes of 29 New York City rats, and compared them with those of rats from northeast China, the presumed ancestral home of brown rats (*Rattus norvegicus*). The researchers looked for genome regions containing variations that were likely to be so beneficial to New York City rats that they quickly became common. The scan produced dozens of such genes, including some associated with diet, behaviour and mobility – perhaps reflecting the challenges, and delights, of life in the Big Apple.

The scientists can't yet say how these genomic hallmarks influence the animals' biology. But future tests in transgenic lab rats could help to explain.

News in focus



The Marine Biological Laboratory in Woods Hole, Massachusetts.

BIOLOGIST EXITS PRESTIGIOUS POST YEARS AFTER VIOLATING SEXUAL-HARASSMENT POLICY

The incident raises important questions about how institutions handle accusations of harassment that occurred at different universities – particularly in the #MeToo era.

By Amy Maxmen

For the past few years, graduate students applying for a prestigious summer course at the Marine Biological Laboratory (MBL) in the harbourside town of Woods Hole, Massachusetts, have been quietly warned about the course's co-director – Richard Schneider. In 2013, an investigation at his institution, the University of California, San Francisco (UCSF), found that he had violated its sexual harassment policy.

Although media reports in 2017 had published some details of Schneider's case,

the situation was discussed only in hushed tones among researchers involved with the MBL embryology course. That changed in mid-January, when a young developmental biologist, Carolyn Dundes, tweeted: "Was super stoked to apply to an MBL course this summer but an ally informed me that the course co-director violated UCSF policy on sexual harassment."

Two days later, Schneider resigned. On 24 January, he was replaced as co-director.

This comes as a relief for some scientists and alumni affiliated with the course who have been uncomfortable ever since Schneider's

violation was made public in 2017 – a few months before the first summer course that he co-directed. (Because directorships last for five years, it was expected he would finish in 2021.) In the past few years, scientists who have participated in the programme have quietly grappled with what to do. Some worried that Schneider might repeat the offence; others felt guilty by association; and some simply wished it had been addressed head on. Dundes found it troubling enough to abandon plans to apply.

"It's horrible – every summer, the students find out," says one instructor, who asked for anonymity to protect against retribution.

Several other scientists who have taught or taken the course spoke to *Nature* on condition of anonymity for the same reason that they didn't speak up earlier: the MBL embryology course is taught by high-ranking biologists who wield significant influence in their fields. Early-career researchers say that speaking up could cost them collaborations, grants or jobs.

The head of the MBL, developmental biologist Nipam Patel, declined to comment on whether he had received complaints about Schneider from students or visiting scientists in previous years. However, he says the MBL has policies barring harassment of all types at the institute.

The public discussion about Schneider and his sudden departure reflect a growing concern about sexual harassment in academia. And they raise important questions about how institutions handle accusations of harassment that occurred at different universities. Many, including the MBL, lack policies about vetting candidates for previous misconduct, which can be especially difficult given that attitudes and discussion about the subject have changed in the past 2.5 years.

"Academic institutions are struggling with how to deal with allegations that predated the #MeToo movement," says Debra Katz, a civil-rights lawyer specializing in sexual-assault and harassment cases at the firm Katz, Marshall & Banks in Washington DC. The hashtag #MeToo went viral in October 2017. And now, at the MBL and elsewhere, Katz says, "Students are responding to the cultural shift, and saying, 'No, we don't want to be in close proximity with someone who has harassed other students in academia.'"

The investigation

The MBL discussion concerns a covert sexual relationship between Schneider and a graduate student, which began weeks after she joined his lab in 2008, at the age of 22. The details of their sexual relationship are described in a report by a committee that investigated a complaint the student filed to UCSF in 2012. UCSF provided a redacted version of the report to *Nature*.

The student, who requested anonymity to protect her from stigmatization, told *Nature* that the physical relationship started when Schneider invited her to a party at UCSF. They drank alcohol, then went to a strip club, where the student says their first sexual encounter happened – and this is substantiated in the investigation report. "At the time, I felt like he valued me scientifically," she recalls. "I felt like this is what a fun scientist would do."

For the next two and a half years, Schneider and the student had a sexual relationship that they kept private. The student says she experienced mounting anxiety over the relationship. "I didn't realize how dependent I was on his approval – what conferences I could go to,

what projects I could work on, my references," she says. "He was my thesis adviser, I couldn't graduate without his approval."

In 2012, she asked for formal mediation because she could no longer work in Schneider's presence. She says that Schneider told her that if others found out about their relationship, it would ruin both of their reputations. Looking back on their relationship, she says, "I don't think it could be called consensual with that kind of power imbalance."

The investigation, which interviewed 13 witnesses, found that "although the relationship may have begun as consensual, the evidence supports a finding that the Complainant, at some point, felt coerced to continue the relationship and reasonably believed that she had no choice but to continue the

"Institutions are struggling with how to deal with allegations that pre-dated the #MeToo movement."

relationship lest it damage her career".

Schneider did not reply to multiple requests for comment. But in the investigation report from UCSF, Schneider "maintains that their relationship was welcome and consensual from beginning to end".

The report concludes that Schneider's "actions and behavior are in violation of the UC Policy on Sexual Harassment". Two years later, in February 2015, UCSF chancellor Sam Hawgood informed Schneider through a letter that he would be disciplined with a demotion from professor to associate professor.

The next year, Schneider won a 'Mentor of the Year' award from UCSF. (The university says he was selected for the prize by students.) He continues to supervise researchers in his UCSF lab. Meanwhile, the student left academia after earning her PhD. "I went into a deep depression," she says to *Nature*. "I had panic attacks and crippling nightmares for years."

Intense environment

Schneider's career continued to advance. In December 2016, the MBL announced that he would co-direct its embryology summer course. During these programmes, around 20 trainees, mainly in their early twenties, live alongside the course directors for six weeks in Woods Hole. Patel says that Schneider's violation wasn't known when he was appointed.

But several people affiliated with the course said they discovered the violation soon afterwards. That's because in early 2017, in response to public-records requests, the University of California gave media outlets more than 100 redacted records on harassment cases across its campuses from 2013 to 2016. *The Mercury News*, a paper based in the San Francisco Bay

area, reported on Schneider's case. In March, two databases on sexual harassment in academia posted his violation online.

By 2019, many graduate students and post-doctoral researchers in the course were aware of Schneider's past because their colleagues had sent them links to the databases and media articles. "I was frankly very frustrated because the embryology course is known to be amazing, so I went but was on guard," says a graduate student who took the course last year, and who asked to remain anonymous to avoid retribution. The student adds, "Sometimes I would imagine the person who almost left grad school because of [Schneider's] actions, and wonder what that person would think."

At least one trainee wasn't bothered. "Rich [Schneider] paid his debt to society, and there are a lot of male scientists who have never been caught," the researcher says on condition of anonymity.

On 14 July 2019, the last day of that year's course, Schneider brought up his violation during an ethics lesson, and apologized if it had made the students uncomfortable, according to a few students present. "I don't think anyone commented," one of them recalls.

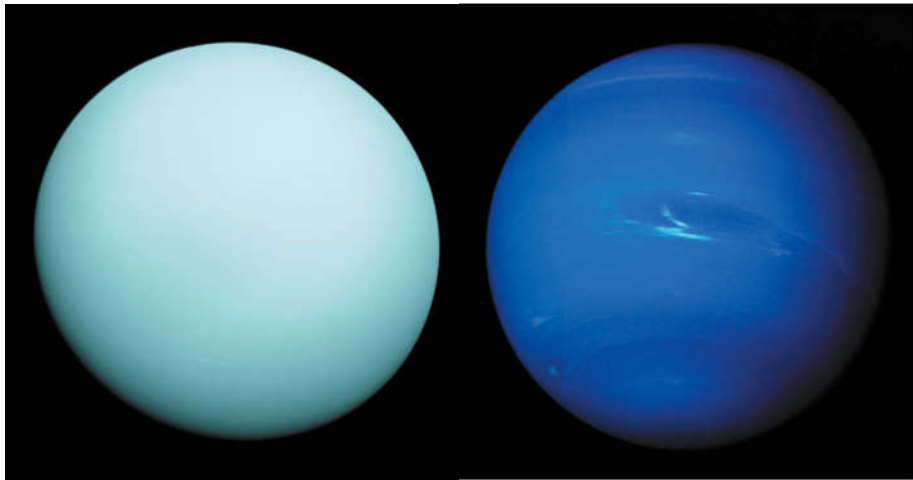
But the situation changed quickly after Dundes's tweet on 14 January. Within 42 hours, more than 14,000 people had seen the tweet, and 824 had clicked on a link that Dundes had posted to an account of Schneider's violation in UCSF's student newspaper, *Synapse*. Mark Peifer, a cell biologist at the University of North Carolina, Chapel Hill, replied with a link to an entry on Schneider in one of the databases of sexual misconduct. "This is really disturbing -- @MBLSscience -- what do you say about this," he wrote.

Patel says the MBL has been developing a plan for how to vet investigators who violated codes of misconduct elsewhere. "Frankly, most institutions are not going to tell us this information," Patel says, "so that is our challenge."

But it's not all that hard, counters Julie Libarkin, a geologist at Michigan State University in East Lansing, who created one of the online databases of substantiated sexual-harassment claims in 2016. Schneider's case and more than 1,000 others are in it. Libarkin acknowledges that her database is incomplete because it includes only records that have been made public – not those that were handled confidentially by institutions.

"A good step would be to require all job candidates to affirm that there has never been a formal or informal finding of misconduct against them," she says. "In order to have a sustainable academic system, we need to put people before everything else," she adds. "These are deep and troubling conversations to have, but they are so important."

Amy Maxmen, a senior reporter at *Nature*, attended the MBL course in 2003.



Uranus (left) and Neptune (right), imaged by Voyager 2, the only probe to have visited them.

RARE CHANCE TO REACH ICE GIANTS EXCITES SCIENTISTS

A planetary alignment provides a favourable window for visiting Uranus and Neptune – but time is tight.

By Elizabeth Gibney

Momentum is building among planetary scientists to send a major mission to Uranus or Neptune – the most distant and least explored planets in the Solar System. Huge gaps remain in scientists' knowledge of the blueish planets, known as the ice giants, which have been visited only once by a space probe. But the pressure is on to organize a mission in the next decade, because scientists want to take advantage of an approaching planetary alignment that would cut travel time.

Interest in the ice giants has grown exponentially, says Amy Simon, a planetary scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland, who co-organized a meeting at the Royal Society in London in January, dedicated to exploring ideas for such a mission. NASA's Voyager 2 is the only spacecraft to have visited Uranus and Neptune, in brief fly-bys in the 1980s. The ice giants therefore represent fresh territory for a wide range of researchers – including the study of planetary rings, atmospheres, moons and oceans, says Simon.

Jovian boost

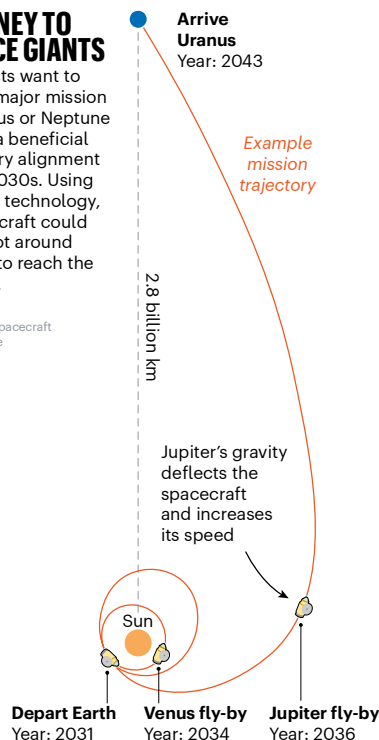
The celestial alignment, between Neptune, Uranus and Jupiter, next occurs in the early 2030s, and would allow a spacecraft to slingshot around Jupiter on its way to the planets. This would reduce the travel time, and allow the craft

to arrive within the lifetimes of its instruments and power systems – usually about 15 years. It would also cut fuel mass, enabling the craft to carry a full suite of scientific instruments (see 'Journey to the ice giants'). To take advantage of the alignment, a mission to Neptune would need to launch by around 2031 and

JOURNEY TO THE ICE GIANTS

Scientists want to send a major mission to Uranus or Neptune during a beneficial planetary alignment in the 2030s. Using existing technology, a spacecraft could slingshot around Jupiter to reach the planets.

Planets & spacecraft not to scale



one to Uranus by the mid-2030s.

The window is “the right time to launch”, Mark Hofstadter, a planetary scientist at the Jet Propulsion Laboratory in Pasadena, California, said at the London meeting. “We don’t want to miss this one.” But the timing is tight. NASA is the most likely space agency to lead the kind of multibillion-dollar ‘flagship’ mission that scientists want. These typically take seven to ten years to prepare, and any green light from NASA would depend on the mission being prioritized in the agency’s Planetary Science Decadal Survey, which reports in 2022. A mission to Neptune or Uranus would also face competition from proposals to return a sample from Mars or explore Venus.

But whereas Mars and Venus scientists are building on decades of exploration, “Uranus and Neptune are genuinely out on their own, as we haven’t completed the very first phase of their exploration yet”, says Leigh Fletcher, a planetary scientist at the University of Leicester, UK, who co-organized the meeting.

Fletcher says that a mission to either planet should include going into orbit around it and sending a probe into its atmosphere or to one of its moons, as the Cassini–Huygens mission did at Saturn.

Blue mysteries

Scientists think of the two planets as twins because of their similar sizes and masses. But no one knows how similar they are, their composition or how they formed, Ravit Helled, a planetary scientist at the University of Zurich, Switzerland, told the meeting. Models struggle to explain the planets’ structures, and why more distant Neptune seems to be warmer than Uranus. It’s assumed that they are made of forms of water, or maybe ammonia ice, said Helled. “But actually we don’t really know that.”

A major mission to the ice giants would also benefit exoplanet studies, said Hannah Wakeford, an exoplanet scientist at the University of Bristol, UK. About 40% of known exoplanets are ice-giant-sized; understanding what these planets’ sizes and atmospheres reveal about their formation relies on understanding those in our own Solar System.

Delegates at the meeting agreed that they would be happy to visit either planet, because both would yield rich results. Studies show that it would be feasible to send probes in a single mission to both planets, but this would be prohibitively expensive. Neptune is appealing because its moon Triton seems to be geologically active and might host a subsurface ocean, potentially of liquid water.

But Uranus – which has a magnetic field that is tilted relative to its rotation axis – has more “odd” features than Neptune does, which challenge existing scientific models, said Hofstadter. The later launch window for Uranus also makes the planet a more realistic target, says Fletcher.

CHINA BANS CASH REWARDS FOR PUBLISHING

New policy tackles perverse incentives that might encourage questionable research practices.

By Smriti Mallapaty

Chinese institutions have been told to stop paying researchers bonuses for publishing in journals, as part of a new national policy to cut perverse incentives that encourage scientists to publish lots of papers rather than focus on high-impact work.

In an order released last week, China's science and education ministries also say that institutions must not promote or recruit researchers solely on the basis of the number of papers they publish, or their citations. Researchers are welcoming the policy, but say that it could reduce the country's competitiveness in science.

In China, one of the main indicators used to evaluate researchers, allocate funding and rank institutions is metrics collected by the Science Citation Index (SCI), a database of articles and citation records for more than 9,000 journals. Since 2009, the number of articles in these journals written by authors from Chinese institutions increased from some 120,000 a year to 450,000 in 2019. Some institutions even pay researchers bonuses for publishing in them.

These practices have incentivized researchers to publish lots of papers at the expense of quality, says Jin Xuan, a chemical engineer at Loughborough University, UK. Evidence suggests that the focus on metrics has also driven a rise in inappropriate practices, such as researchers submitting plagiarized or fraudulent papers, or inappropriately citing their own or a colleague's work to boost citations (L. Tang *et al.* *J. Assoc. Inf. Sci. Tech.* **66**, 1923–1932; 2015).

The goal of the new policy is not to discourage Chinese researchers from publishing papers in SCI-listed journals, but to stop inappropriate publishing and citation practices, says Tang Li, a researcher of science and technology policy at Fudan University in Shanghai, China.

Xuan adds that the policy aligns well with global declarations, such as the San Francisco Declaration on Research Assessment, that aim to move away from an over-reliance on these types of metric in research appraisals and to limit perverse incentives that drive researchers to engage in questionable research practices.

As part of the new policy, researcher assessments will now need to use indicators of the quality of research, such as how innovative the work is, and whether it represents a significant scientific advance or contributes to solving important societal problems. These evaluations should also rely more heavily on the professional opinions of expert peers, and consider research in journals published in China, many of which are not listed in the SCI.

But Futao Huang, who studies higher-education policy at Hiroshima University, Japan, says it is not clear what exactly the new evaluation system will look like, because

the ministry's notices lack specific, practical recommendations.

Huang thinks the new measures could result in a drop in the number of low-quality or fraudulent papers, but might also trigger a decline in China's total publications in indexed journals as researchers feel less pressure to publish to gain degrees, promotions or funding.

And fewer Chinese papers in indexed journals could affect the country's research competitiveness, says Huang. International researchers might be less inclined to collaborate with Chinese academics without a publication record in these journals, and fewer papers could push Chinese universities lower down in international rankings, he says.

Xuan says the focus on assessing researchers on the basis of their work in Chinese journals is controversial because a lot of them publish in Mandarin, and the journals are unknown to scientists outside China.

Other scientists have raised concerns about the new assessments relying too heavily on peer reviews, which are subjective and could create conflicts of interest or place too much emphasis on personal relationships.

MYSTERY DEEPENS OVER ANIMAL SOURCE OF CORONAVIRUS

Pangolins are a prime suspect, but a slew of genetic analyses has yet to find conclusive proof.

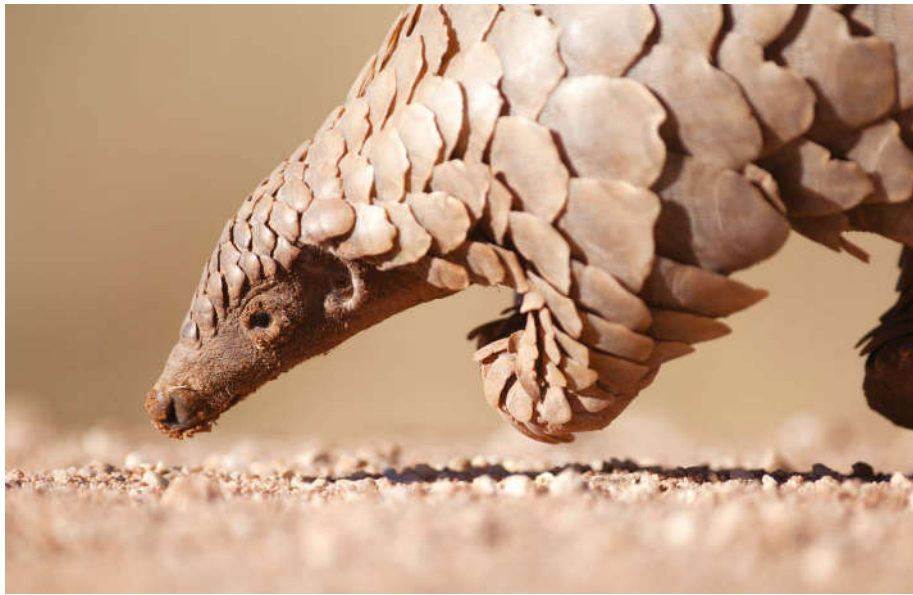
By David Cyranoski

Scientists are racing to identify the source of the coronavirus causing havoc around the world. Last month, Chinese researchers suggested, on the basis of genetic analyses, that the scaly, ant-eating pangolin was the prime suspect. But scientists have now examined those data – along with three similar genome studies – and say that although the mammal is still a contender, the mystery is far from solved.

Health officials want to pin down the virus's source so they can prevent new outbreaks. Scientists assume that the pathogen jumped to people from an animal, as other coronaviruses have; for example, the virus that causes severe acute respiratory syndrome (SARS) is thought to have jumped to humans from civets in 2002. Dozens of people infected early in the current outbreak worked at a live-animal market in the Chinese city of Wuhan, but tests of coronavirus samples found at the market have yet to identify a source.

Three separate Chinese teams are trying to trace the origin of the coronavirus. Researchers at the South China Agricultural University in Guangzhou suggested pangolins as the animal source at a press conference on 7 February. Pangolins are sought-after in China for their meat and scales. Although the animals can't be sold in China owing to a worldwide ban, they are still smuggled in from elsewhere in Asia and Africa. The researchers said they had found a coronavirus in smuggled pangolins that was a 99% genetic match to the human virus.

But the result did not pertain to the entire genome. In fact, it related to a specific site known as the receptor-binding domain (RBD), say the study's authors, who posted their analysis¹ on the preprint server bioRxiv on 20 February. The press-conference report was the result of an “embarrassing miscommunication between the bioinformatics group and the lab group of the study”, explains Xiao Lihua, a parasitologist at the South China Agricultural University and a co-author of the paper. A whole-genome comparison found that the



Pangolins are often smuggled into China, where there is demand for their meat and scales.

pangolin and human viruses share 90.3% of their DNA.

The RBD is a crucial part of coronaviruses that allows them to latch on to and enter a cell. Even a 99% similarity between two viruses' RBDs is not necessarily enough to link them, says Linfa Wang, a virologist at Duke–National University of Singapore Medical School.

Three similar comparison studies were also posted on bioRxiv last month. One, posted on 18 February, found² that coronaviruses in frozen cell samples from illegally trafficked pangolins shared between 85.5% and 92.4% of their DNA with the virus found in humans.

Two other papers^{3,4} published on 20 February studied coronaviruses from smuggled

pangolins. These showed 90.23% and 91.02% similarity, respectively, to the new coronavirus.

Higher genetic similarity is needed before the host can be definitively identified, says Arinjay Banerjee, who studies coronaviruses at McMaster University in Hamilton, Canada. He notes that the SARS virus shared 99.8% of its genome with a civet coronavirus.

So far, the closest match to the new coronavirus has been found in a bat in China's Yunnan province. A study⁵ published on 3 February found that the bat coronavirus shared 96% of its genetic material with the virus that causes COVID-19. Bats could have passed the virus to people, but scientists think it was probably transmitted through an intermediate host.

But if pangolins are a host, and they came from another country, it raises the question of why there haven't been reports of people being infected there, asks Jiang Zhigang, an ecologist at the Chinese Academy of Sciences Institute of Zoology in Beijing.

1. Xiao, K. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/2020.02.17.951335> (2020).
2. Lam, T. T.-Y. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/2020.02.13.945485> (2020).
3. Liu, P. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/2020.02.18.954628> (2020).
4. Zhang, T., Wu, Q. & Zhang, Z. Preprint at bioRxiv <https://doi.org/10.1101/2020.02.19.950253> (2020).
5. Zhou, P. *et al.* *Nature* <https://doi.org/10.1038/s41586-020-2012-7> (2020).

Fundación BBVA

13th Edition BBVA Foundation Frontiers of Knowledge Awards

With the collaboration of



The **BBVA Foundation Frontiers of Knowledge Awards** recognize and reward world-class research and artistic creation, prizing contributions of singular impact for their originality and significance. The name of the scheme is intended to denote not only research work that substantially enlarges the scope of our current knowledge – pushing forward the frontiers of the known world – but also the meeting and overlap of different disciplinary areas and the emergence of new fields.

The Frontiers of Knowledge Awards honor fundamental disciplinary or interdisciplinary advances across a broad expanse of the knowledge map of the 21st century.

The BBVA Foundation is assisted in the award process by the **Spanish National Research Council (CSIC)**.

Fundación BBVA
Plaza de San Nicolás, 4 · 48005 Bilbao · Spain
Paseo de Recoletos, 10 · 28001 Madrid · Spain
awards-info@bbva.es

Categories

- 1/ **Basic Sciences** (Physics, Chemistry, Mathematics)
- 2/ **Biology and Biomedicine**
- 3/ **Information and Communication Technologies**
- 4/ **Ecology and Conservation Biology**
- 5/ **Climate Change**
- 6/ **Economics, Finance and Management**
- 7/ **Humanities and Social Sciences**
- 8/ **Music and Opera**

In Humanities and Social Sciences, the award will alternate annually between these two disciplinary domains, with this thirteenth edition dedicated to the Humanities.

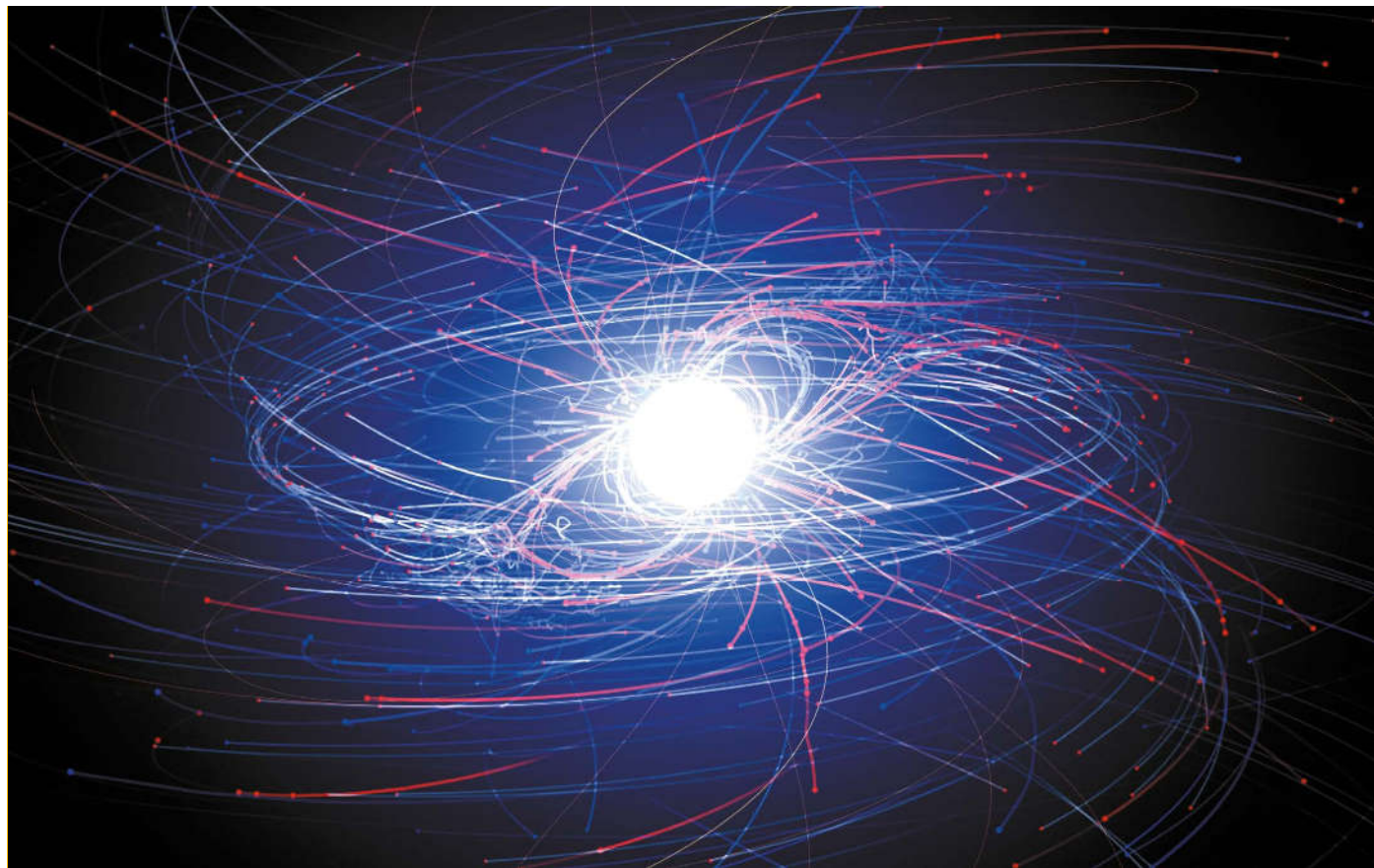
Nomination

Nominations are invited from scientific or artistic societies and organizations, public or private R&D centers, university and hospital departments, schools of music, orchestras, and organizations working on or around the issue of climate change, as well as other institutions specified in the call conditions.

Entry submission

The nomination period concludes at **23:00 GMT on June 30, 2020**.

www.frontiersofknowledgeawards-fbbva.es



NASA'S GODDARD SPACE FLIGHT CENTER

Powerful magnetic and electric fields whip charged particles around, in a computer simulation of a spinning neutron star.

THE STRANGE HEARTS OF NEUTRON STARS

Space observations are poised to reveal more about the centre of one of the Universe's most enigmatic objects. **By Adam Mann**

When a massive star dies in a supernova, the explosion is only the beginning of the end. Most of the stellar matter is thrown far and wide, but the star's iron-filled heart remains behind. This core packs as much mass as two Suns and quickly shrinks to a sphere that would span the length of Manhattan. Crushing internal pressure – enough to squeeze Mount Everest to the size of a sugar cube – fuses subatomic protons and electrons into neutrons.

Astronomers know that much about how neutron stars are born. Yet exactly what happens afterwards, inside these ultra-dense

cores, remains a mystery. Some researchers theorize that neutrons might dominate all the way down to the centre. Others hypothesize that the incredible pressure compacts the material into more exotic particles or states that squish and deform in unusual ways.

Now, after decades of speculation, researchers are getting closer to solving the enigma, in part thanks to an instrument on the International Space Station called the Neutron Star Interior Composition Explorer (NICER).

Last December, this NASA space observatory provided astronomers with some of the most precise measurements ever made of a neutron star's mass and radius^{1,2}, as well as unexpected findings about its magnetic field^{1,3}. The NICER

team plans to release results about more stars in the next few months. Other data are coming in from gravitational-wave observatories, which can watch neutron stars contort as they crash together. With these combined observations, researchers are poised to zero in on what fills the innards of a neutron star.

For many in the field, these results mark a turning point in the study of some of the Universe's most bewildering objects. "This is beginning to be a golden age of neutron-star physics," says Jürgen Schaffner-Bielich, a theoretical physicist at Goethe University in Frankfurt, Germany.

Launched in 2017 aboard a SpaceX Falcon 9 rocket, the US\$62-million NICER telescope

sits outside the space station and collects X-rays coming from pulsars – spinning neutron stars that radiate charged particles and energy in enormous columns that sweep around like beams from a lighthouse. The X-rays originate from million-degree hotspots on a pulsar's surface, where a powerful magnetic field rips charged particles off the exterior and slams them back down at the opposing magnetic pole.

NICER detects these X-rays using 56 gold-coated telescopes, and time-stamps their arrival to within 100 nanoseconds. With this capability, researchers can precisely track hotspots as a neutron star whips around at up to 1,000 times per second. Hotspots are visible as they swing across the object. But neutron stars warp space-time so strongly that NICER also detects light from hotspots facing away from Earth. Einstein's general theory of relativity provides a way to calculate a star's mass-to-radius ratio through the amount of light-bending. That and other observations allow astrophysicists to pin down the masses and radii of the deceased stars. Those two properties could help in determining what is happening down in the cores.

Deep, dark mystery

Neutron stars get more complicated the deeper one goes. Beneath a thin atmosphere made mostly of hydrogen and helium, the stellar remnants are thought to boast an outer crust just a centimetre or two thick that contains atomic nuclei and free-roaming electrons. Researchers think that the ionized elements become packed together in the next layer, creating a lattice in the inner crust. Even further down, the pressure is so intense that almost all the protons combine with electrons to turn into neutrons, but what occurs beyond that is murky at best (see 'Dense matter').

"It's one thing to know the ingredients," says Jocelyn Read, an astrophysicist at California State University, Fullerton. "It's another to understand the recipe, and how those ingredients are going to interact with each other."

Physicists have some idea of what happens, thanks to particle accelerators on Earth. At facilities such as Brookhaven National Laboratory in Upton, New York, and CERN's Large Hadron Collider near Geneva, Switzerland, researchers have smashed together heavy ions, such as those of lead and gold, to create brief collections of monumentally dense material. But these kinetic experiments generate billion- or even trillion-degree flashes, in which protons and neutrons dissolve into a soup of their constituent quarks and gluons. Terrestrial instruments have a hard time probing the relatively mild millions-of-degrees conditions inside neutron stars.

There are multiple ideas about what might occur. It could be that quarks and gluons roam freely. Or, the extreme energies could lead

to the creation of particles called hyperons. Like neutrons, these particles contain three quarks. But whereas neutrons contain the most basic and lowest-energy quarks, known as up and down quarks, a hyperon has at least one of those replaced with an exotic 'strange' quark. Another possibility is that the centre of a neutron star is a Bose–Einstein condensate, a state of matter in which all subatomic particles act as a single quantum-mechanical entity. And theorists have dreamt up even more outlandish prospects, too.

"It's one thing to know the ingredients. It's another to understand the recipe."

Crucially, each possibility would push back in a characteristic way against a neutron star's colossal gravity. They would generate different internal pressures and therefore a larger or smaller radius for a given mass. A neutron star with a Bose–Einstein condensate centre, for instance, is likely to have a smaller radius than one made from ordinary material such as neutrons. One with a core made of pliable hyperon matter could have a smaller radius still.

"The types of particles and the forces between them affect how soft or squashy the material is," says Anna Watts, a NICER team member at the University of Amsterdam.

Differentiating between the models will require precise measurements of the size and mass of neutron stars, but researchers haven't yet been able to push their techniques to fine-enough levels to say which possibility is most likely. They typically estimate masses by observing neutron stars in binary pairs. As the objects orbit one another, they tug gravitationally on each other, and astronomers can use this to determine their masses. Roughly 35 stars have had their masses measured in this way, although the figures can contain error bars of up to one solar mass. A mere dozen or so have also had their radii calculated, but in many cases, the techniques can't determine this value to better than a few kilometres – as much as one-fifth of the size of a neutron star.

NICER's hotspot method has been used by the European Space Agency's XMM-Newton X-ray observatory, which launched in 1999 and is still in operation. NICER is four times more sensitive and has hundreds of times better time resolution than the XMM-Newton. Over the next two to three years, the team expects to be able to use NICER to work out the masses and radii of another half a dozen targets, pinning down their radii to within half a kilometre. With this precision, the group will be well placed to begin plotting out what is known as the neutron-star equation of state, which

relates mass to radius or, equivalently, internal pressure to density.

If scientists are particularly lucky and nature happens to serve up especially good data, NICER might help eliminate certain versions of this equation. But most physicists think that, on its own, the observatory will probably narrow down rather than completely rule out models of what happens in the mysterious objects' cores.

"This would still be a huge advance on where we are now," says Watts.

Field lines

NICER's first target was J0030+0451, an isolated pulsar that spins roughly 200 times per second and is 337 parsecs (1,100 light years) from Earth, in the constellation Pisces.

Two groups – one based primarily at the University of Amsterdam¹ and another led by researchers at the University of Maryland in College Park² – separately sifted through 850 hours of observations, serving as checks on one another.

Because the hotspot light curves are so complex, the groups needed supercomputers to model various configurations and work out which ones best fit the data. But both came up with similar results, finding that J0030 has a mass that is 1.3 or 1.4 times that of the Sun, and a radius of roughly 13 kilometres.

Those results are not definitive: they could be used to support either the mundane or the otherworldly predictions for what's inside the guts of neutron stars. "There's no requirement for anything funky or crazy or exotic yet," says Andrew Steiner, a nuclear astrophysicist at the University of Tennessee, Knoxville.

Researchers got a bigger surprise with findings about the shape and position of the hotspots. The canonical view of neutron stars has their magnetic field lines looking like those surrounding a bar magnet, with north and south sides emerging from circular spots at opposing ends of the star. By contrast, the Dutch supercomputer simulations implied that both of J0030's hotspots are in its southern hemisphere, and that one of them is long and crescent-shaped¹. The Maryland team also came up with the possibility of a three-hotspot solution: two southerly oval-shaped ones and a final circle near the rotational south pole³.

"It looks like they might have made the first real detection of a pulsar where the beams are not 180 degrees separated," says Natalie Webb, an astrophysicist at the Institute for Research in Astrophysics and Planetology in Toulouse, France, who has modelled such possibilities. "That's fantastic if true."

The results would bolster previous observations and theories suggesting that neutron stars' magnetic fields, which are one trillion times stronger than the Sun's, can be more complex than generally assumed. After they first form, pulsars are thought to slow their

Feature

rotation over millions of years. But if they have a companion star orbiting around them, they might steal material and angular momentum from this partner, boosting their spinning to superfast speeds. As the matter gets deposited on the star's exterior, some theorists suggest it could affect a fluid-like layer of subsurface neutrons, generating gigantic vortices that twist the neutron star's magnetic field into odd arrangements. The companion might ultimately be consumed or lose so much mass that it becomes gravitationally unbound and flies away, as could have been the case with the now-solitary J0030.

Work in progress

NICER is continuing to observe J0030 to further improve the precision of its radius measurements. At the same time, the team is beginning to analyse data from a second target, a slightly heavier pulsar with a white-dwarf companion. Other astronomers have used observations of this pair's orbital dance to determine the pulsar's mass, which means NICER researchers have an independent measurement that they can use to validate their findings.

Among NICER's targets, the team plans to include at least a couple of high-mass pulsars, including the current record-holder for most massive neutron star — a behemoth with a mass 2.14 times that of the Sun. That should allow the researchers to probe an upper limit: the point at which a neutron star collapses into a black hole. Even the 2.14-solar-mass object is challenging for theorists to explain. Several researchers have also suggested that NICER might be able to find two neutron stars with the same mass but different radii. That would suggest the presence of a transition point, at which slight differences create two distinct cores. One might contain mostly neutrons, for example, and the other might be composed of more-exotic material.

Although NICER is at the vanguard, it is not the only instrument plumbing pulsars' depths. In 2017, the US Laser Interferometer Gravitational-Wave Observatory (LIGO), along with the Virgo detector in Italy, picked up the signal from two neutron stars crashing and merging together⁴. As the objects rotated around one another before the crash, they emitted gravitational waves that contained information about the stars' size and structure. Each star's colossal gravitational influence tugged on and deformed its partner, contorting both from spheres into teardrop shapes. The amount of distortion in those final moments gives physicists clues about the malleability of the material inside the neutron stars.

LIGO's facility in Livingston, Louisiana, picked up a second neutron-star smash-up last April, and more events could be spotted at any time. So far, the two mergers have only hinted at the properties of neutron-star

DENSE MATTER

Neutron stars get denser with depth. Although researchers have a good sense of the composition of the outer layers, the ultra-dense inner core remains a mystery.

Outer crust
Atomic nuclei and free electrons

Inner crust
Free neutrons and electrons, heavier atomic nuclei

Outer core
Neutron-rich quantum liquid

Inner core
Unknown, ultra-dense matter

Atmosphere
Mostly hydrogen and helium

Core scenarios

A number of possibilities have been suggested for the inner core, including these three options.

u Up quark s Strange quark
d Down quark \bar{d} Anti-down quark



Quarks

The constituents of protons and neutrons — up and down quarks — roam freely.



Bose-Einstein condensate

Particles such as pions containing an up quark and an anti-down quark combine to form a single quantum-mechanical entity.



Hyperons

Particles called hyperons form. Like protons and neutrons, they contain three quarks but include 'strange' quarks.

interiors, suggesting that they are not particularly deformable. But the current generation of facilities can't observe the crucial final moments, when the warping would be greatest and would display internal conditions most clearly.

The Kamioka Gravitational Wave Detector in Hida, Japan, is expected to come online later this year, and the Indian Initiative in Gravitational-wave Observations near Aundha Naganath, Marathwada, in 2024. In combination with LIGO and Virgo, they will improve sensitivity, potentially even capturing the details of the moments leading up to a crash.

Looking further into the future, several planned instruments could make observations that elude NICER and current gravitational-wave observatories. A Chinese-European satellite called the enhanced X-ray Timing and Polarimetry mission, or eXTP, is expected to launch in 2027 and study both isolated and binary neutron stars to help determine their equation of state. Researchers have also proposed a space-based mission that could fly in the 2030s called the Spectroscopic Time-Resolving Observatory for Broadband Energy X-rays, or STROBE-X. It would use

NICER's hotspot technique, pinning down the masses and radii of at least 20 more neutron stars with even more precision.

The hearts of neutron stars will probably always retain some secrets. But physicists now seem well placed to begin peeling back the layers. Read, who is a member of the LIGO team, says that she has collaborated on a project to imagine what scientific questions gravitational-wave detectors would be able to tackle in the 2030s and 2040s. In the process, she realized that the landscape for neutron-star research — in particular, the question of the equation of state — should look very different by then.

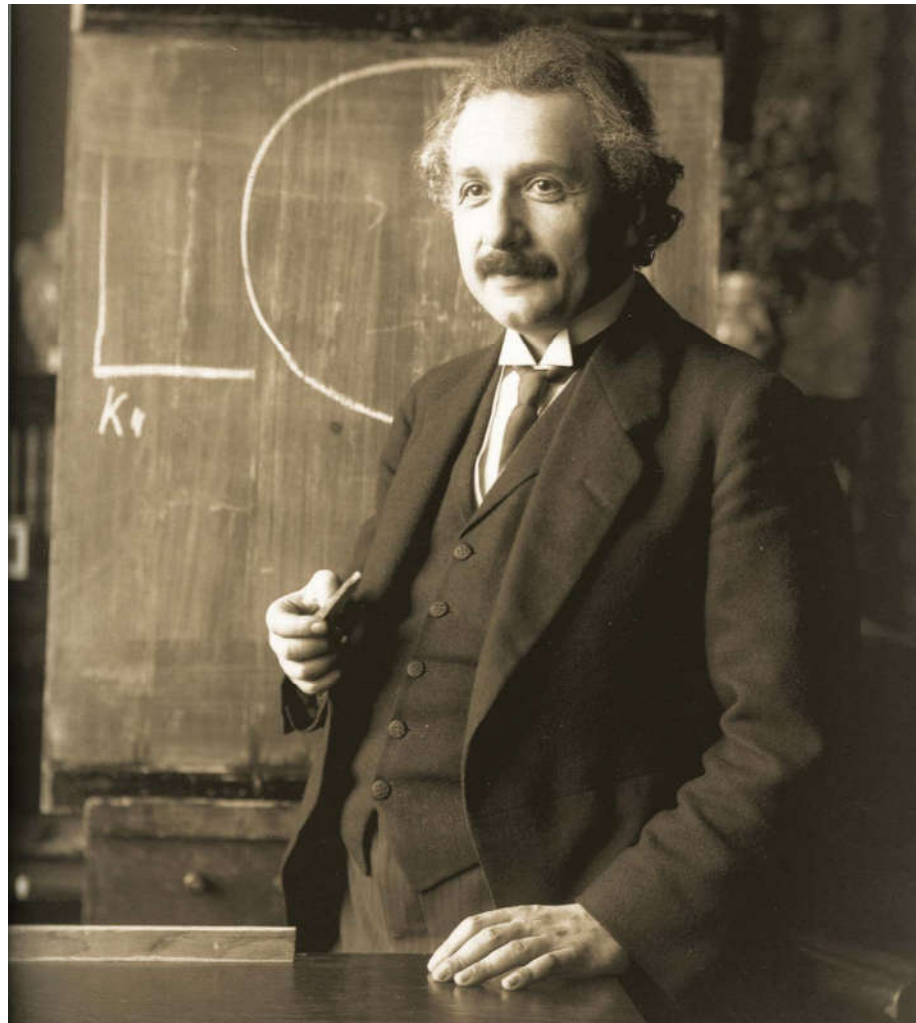
"It's been this long-standing puzzle that you figure will always be there," she says. "Now we're at a point where I can see the scientific community figuring out the neutron-star-structure puzzle within this decade."

Adam Mann is a freelance journalist based in Oakland, California.

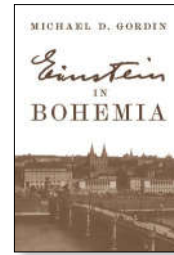
1. Riley, T. E. et al. *Astrophys. J. Lett.* **887**, L21 (2019).
2. Miller, M. C. et al. *Astrophys. J. Lett.* **887**, L24 (2019).
3. Bilous, A. V. et al. *Astrophys. J. Lett.* **887**, L23 (2019).
4. Abbott, B. P. et al. *Phys. Rev. Lett.* **119**, 161101 (2017).

SOURCE: ADAPTED FROM NASA GODDARD SVS

Books & arts



Albert Einstein worked at a university in Prague in 1911–12.



Einstein in Bohemia
Michael D. Gordin
Princeton Univ. Press
(2020)

on developing his general theory of relativity. It was in Prague that Einstein came up with the idea of gravitational lensing, the concept that the pull of stars, planets and other astronomical objects would distort light rays. That idea – tool – is now central to modern astronomy, used to determine, for example, how much dark matter is hovering around clusters of galaxies. Although Einstein had yet to conjure up a dynamical theory of space-time, it was his prediction of gravitational lensing that would demonstrate the theory's chops in 1919, when physicist Arthur Eddington and his colleagues measured how light from stars in the Hyades cluster was deflected by the gravitational pull of the Sun during a solar eclipse.

Personal impact

At a personal level, gaining the position of professor of theoretical physics at the German University in Prague pushed Einstein into the senior echelons of academia. Just three years before, he had been a patent clerk in Bern. Interestingly, Gordin tells us, he was not the first person to be offered the job: first refusal went to one Gustav Jaumann at the German Technical University in Brno (who remembers him?), who turned it down.

It was also in Prague that Einstein's marriage to fellow physicist Mileva Marić began to fall apart. She was miserable there: snubbed as a Serbian and resentful at being dragged around for her husband's career, then left sitting at home while he travelled elsewhere to give talks and collaborate. It was during a trip to Berlin that he began an affair with his cousin Elsa Löwenthal, who eventually became his second wife.

But that is about it. Slim pickings on which to base a whole book, I feared. Yet Gordin does something ingenious. He uses Einstein as a MacGuffin, a device that propels the plot but has little significance to the story he wants to tell. He explodes the narrative out of what he calls the "spacetime interval" of 1911–12 to follow a host of figures who were involved with Einstein in Prague, in some cases very tangentially. In so doing, he careers through the history of ideas as well as the political

A fresh look at Einstein's Prague period

Michael Gordin's elegant prose uses 16 months to build a panoramic view of a century. **By Pedro Ferreira.**

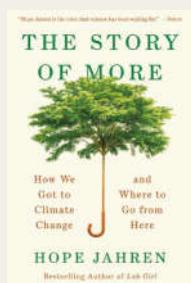
Many people skip over the fact that, from early April 1911 to late July 1912, Albert Einstein lived in Prague. "It was, after all, such a short time, and quite early in the physicist's career," Michael Gordin explains at the start of *Einstein in Bohemia*. Historians have variously dismissed those 16 months as an interlude, a

sojourn and a detour.

So did I – before I read this improbably good book.

Multiple biographies of Einstein agree that the Prague period is notable for one main reason. Free of the heavy teaching load that had burdened him as an associate professor at the University of Zurich in Switzerland, he focused

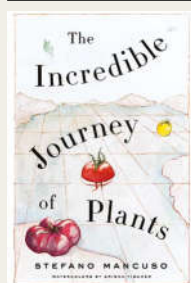
Books in brief



The Story of More

Hope Jahren Vintage (2020)

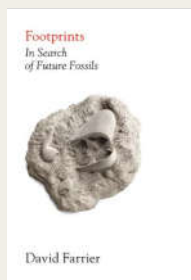
In 2009, palaeobiologist Hope Jahren was required to teach climate change. Initially reluctant, she soon conceived a vocation. Her compelling book uses statistics brilliantly to provoke self-examination. In sections on 'Life', 'Food', 'Energy' and 'Earth', it illuminates subjects from population growth to melting glaciers. If the whole planet consumed resources on the US scale, carbon dioxide emissions would be more than four times higher, she observes: "Using less and sharing more is the biggest challenge our generation will ever face."



The Incredible Journey of Plants

Stefano Mancuso (transl. Gregory Conti) Other Press (2020)

About 400 metres from ground zero in Hiroshima, a weeping willow and other plants regrew from their roots. Revered, they are labelled *hibakujumoku*, "trees that suffered an atomic explosion", an elderly Japanese diplomat translates in flawless Italian for visiting plant neurobiologist Stefano Mancuso. Later, he confesses he is a *hibakusha*: he survived the strike because his classroom was protected by a curtain of trees. Such anecdotes enliven Mancuso's quirky little global history, which argues that plants "are more sensitive than animals".



Footprints: In Search of Future Fossils

David Farrier Farrar, Straus and Giroux (2020)

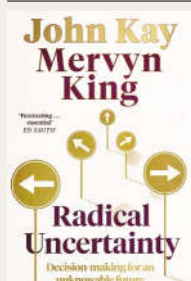
Fossil footprints unmasked by a 2013 storm on the English coast revealed that hominins walked beside an estuary 850,000 years ago. Although quickly erased by the tide, they inspired David Farrier to consider modern civilization's future footprints, including Neil Armstrong's marks on the Moon and the nuclear footprint: a geological repository for Finland's spent fuel. This is designed to be forgotten — unlike its US equivalent, which proposes to use warning signs modelled on Edvard Munch's 1893 painting *The Scream*.



The Future of Brain Repair

Jack Price MIT Press (2020)

In 1996, neurobiologist Jack Price, then at a major pharmaceutical company, was invited to fund academic research into stem-cell therapies. He declined. Now an academic himself, he is more hopeful. In 2006, Shinya Yamanaka discovered how to make 'pluripotent' stem cells, enabling brain-like tissue to be generated in a dish — "albeit small, misshapen and underdeveloped", as Price notes in his clear, honest but intellectually challenging account. Today, several therapies have entered clinical trials. But how to make them affordable?



Radical Uncertainty

John Kay and Mervyn King Bridge Street (2020)

When Christopher Columbus sought a westerly route to the Indies, "whatever counted as cost-benefit analysis in the Spanish court took no account of the possibility of a New World", say economists John Kay and Mervyn King. They refreshingly criticize their discipline for not recognizing that its use of 'risk', 'uncertainty' and 'rationality' doesn't match that of lay people. Odd, then, that their far-ranging book on "radical uncertainty" mentions Max Planck's dalliance with economics but not Werner Heisenberg's uncertainty principle. **Andrew Robinson**

turmoil of Bohemia (now part of the Czech Republic) during most of the twentieth century, touching on physics, philosophy, nationhood, anti-Semitism and the rise of Prague as a centre of intellectual life.

There are quirky observations, almost worthy of playwright Tom Stoppard. For example, Einstein and writer Franz Kafka probably met at a 1911 cultural soirée in the house of Berta Fanta, a "philosophically ambitious" socialite who held a salon above her husband's pharmacy in Prague's Old Town Square.

Social circle

But what really grips are the people. Take Oskar Kraus, a philosopher at the German University. Originally trained in law, he took against Einstein, writing countless articles in philosophy journals unpicking what he saw as egregious internal inconsistencies in relativity. His writing and stance foreshadowed the anti-relativity strand of the Deutsche Physik movement, an eviscerating force in German academia during the rise of the Third Reich. Kraus, who had been born into a Jewish family but converted to Protestantism, was arrested by the Gestapo and ultimately fled to Oxford, UK.

Inevitably, Gordin takes in Ernst Mach, who had been in a post similar to Einstein's at the German University's forerunner from 1867 to 1895. Mach had been "the most successful physicist in the university's history" and played an important role as rector for part of his tenure. But, like Einstein, he had been the second choice for the post. Mach's ideas shaped the work of important relativists after Einstein, such as Dennis Sciama and Robert Dicke.

Another pen portrait is of Einstein's successor in the post, physicist and philosopher Philipp Frank. His journey through the turbulent Prague of the 1930s serves as spotlight on a place battered by historical forces. During the late 1920s and early 1930s, Frank was part of the Vienna Circle, a hugely influential group of scientists and philosophers that also included philosopher Rudolf Carnap and mathematician Kurt Gödel. In Prague, Frank did much to carry the flame of both Einstein and Mach's ideas through books and journal articles, publicly sparring with Kraus whenever necessary. In 1938, he had to flee to the United States, where he ended up at Harvard University in Cambridge, Massachusetts, and wrote one of Einstein's first and most notable biographies.

This is a panoramic view of twentieth-century Bohemia, with a sprinkling of Einstein. But what really carries it through is the beauty and force of Gordin's prose.

Pedro Ferreira is professor of astrophysics at the University of Oxford, UK, and author of *The Perfect Theory*.
e-mail: pedro.ferreira@physics.ox.ac.uk

Comment



KUNI TAKAHASHI/GETTY

Women carry coal from an open-cast mine in Jharkhand state in India.

Emissions: world has four times the work or one-third of the time

Niklas Höhne, Michel den Elzen, Joeri Rogelj, Bert Metz, Taryn Fransen, Takeshi Kuramochi, Anne Olhoff, Joseph Alcamo, Harald Winkler, Sha Fu, Michiel Schaeffer, Roberto Schaeffer, Glen P. Peters, Simon Maxwell & Navroz K. Dubash

New synthesis shows what a wasted decade means for the climate pact made in Paris.

The past decade of political failure on climate change has cost us all dear. It has shrunk the time left for action by two-thirds. In 2010, the world thought it had 30 years to halve global emissions of greenhouse gases. Today, we know that this must happen in ten years to minimize the effects of climate change. Incremental shifts

that might once have been sufficient are no longer enough.

The further bad news is that, even taken together, the proposed climate action by all countries is a long way from meeting this requirement. Rather than halving emissions by 2030, countries' climate proposals will lead to a slight increase. Worse still, individual countries

Comment

are not on track to achieve commitments that were insufficient from the outset and are now woefully inadequate.

The better news is that more countries, regions, cities and businesses are implementing the deep, rapid transformations that are urgently required. At scale, these could achieve the collective climate goals that nations agreed in Paris more than four years ago. There are lessons to be learnt from places such as Costa Rica, Shenzhen in China and Copenhagen that have made strides through the use of renewable energy and electrified transport. The United Kingdom (together with 75 other parties) and California have at least set ambitious goals to become carbon neutral, which might send signals to industry even before supporting policies are implemented. Meanwhile, 26 banks have stopped directly financing new coal-fired power plants (see go.nature.com/32uped2).

Much is happening on the ground. The question is how to ramp up these activities fast enough to keep warming to less than 1.5 °C above pre-industrial levels.

Here we present a snapshot of the extent to which nations' individual pledges are inconsistent with their stated collective goals. We also note some of the pockets of promise. We draw our conclusions from a synthesis of all ten editions of the *Emissions Gap Report* produced by the United Nations Environment Programme (UNEP)^{1–5}. Each year for the past decade, this report has examined the difference between what countries have pledged to do individually to reduce greenhouse-gas emissions, and what they need to do collectively to meet agreed temperature goals – the 'gap'.

Our analysis shows that the gap has widened by as much as four times since 2010. There are three reasons for this. First, global annual greenhouse-gas emissions increased by 14% between 2008 and 2018 (ref. 6). This means that



Electric taxis at a charging station in Shenzhen, China.

QILAI SHEN/BLOOMBERG/GETTY

emissions now have to decline faster than was previously estimated, because it is cumulative emissions that determine the long-term temperature increase. Second, the international community now agrees that it must ensure a lower global temperature rise than it decided ten years ago, because climate risks are better understood. And third, countries' new climate

pledges have been insufficient.

The tenth anniversary of the report coincides with the 2020 milestone to which countries agreed in Paris. They undertook to communicate or update climate pledges, or 'nationally determined contributions', to the UN Framework Convention on Climate Change conference (COP26) this November in Glasgow, UK. Clearly, the promises must be overhauled – and then, crucially, kept – if the yawning gap between 'talk and walk' is going to close by 2030.

Gap minder

The scope of the UNEP emissions gap reports has evolved over time, in line with climate policy. So what has changed during the past decade?

In the 2009 Copenhagen accord⁷ and the 2010 Cancun agreement⁸, countries collectively pledged to limit warming to below 2 °C, and 73 countries individually pledged emissions targets for 2020. The 2015 Paris agreement, responding to mounting concern over the impacts of climate change, tightened the collective temperature limit to "well below 2 °C" and agreed "to pursue efforts to limit the temperature increase to 1.5 °C" (ref. 9). Under the Paris deal, 192 parties individually pledged emissions targets, typically for 2030 (see 'More and faster').

From 2010 to 2014, the gap reports projected

THE SEVEN TOP EMITTERS

Country or region (2018 emissions in gigatonnes CO ₂ equivalent) ^{6,13}	Change in projected greenhouse-gas emissions by 2030 since 2015	Potential reasons
China (13.2)	No change	New climate and energy policies; altered growth projections.
United States (6.6)	No change	Rollback of federal policies works against price drops in renewables and reductions in coal use.
European Union (4.0)	Lower	Mostly attributable to implementation of new policies.
India (3.8)	Slightly lower	Unclear.
Russia (2.4)	No change	No change in policies or growth projections.
Indonesia (2.3)	Higher	Higher emissions projections from deforestation.
Brazil (1.6)	Higher	Higher emissions projections from deforestation.

Comparison of the 2015 and 2019 UNEP *Emissions Gap Report*^{2,20} and other sources^{21–25} provides information about changes in current policy projections for the leading emitters. Uncertainties for each estimate are large. See Supplementary Information for details.

the likely difference in 2020 between the expected result of countries' pledges and the pathways towards 2 °C. The 2010 report documented a shortfall of 14%. Since 2015, the reports have forecast the expected shortfall in 2030 between the countries' pledges and progress towards both 1.5 °C (current shortfall of 55%) and 2 °C (current shortfall of 25%; see 'More and faster'). The report also examines the policies that countries are implementing domestically.

Had serious climate action begun in 2010, the cuts required to meet the emissions levels for 2 °C would have been around 2% per year, on average, up to 2030. Instead, emissions increased. Consequently, the required cuts from 2020 are now more than 7% per year on average for 1.5 °C (close to 3% for 2 °C).

The time window for halving global emissions has also narrowed drastically. In 2010, it was 30 years; today, it is 10 years for 1.5 °C (25 years for 2 °C). Although many reports, scientists and policymakers continue to discuss rises of 2 °C, it must be emphasized that, in 2018, the Intergovernmental Panel on Climate Change reported that warming of more than 1.5 °C would be disastrous¹⁰.

Countries are not even on track to achieve their now plainly inadequate 2015 pledges. Of the G20 countries, seven (Australia, Brazil, Canada, Japan, South Korea, South Africa and the United States) need to implement existing policy or roll out new measures. (The United States has begun the process of withdrawing from the Paris agreement, and will leave in November.) Russia and Turkey have set themselves unambitious targets that they can meet without new policies.

Since 2015, estimated global emissions in 2030 have decreased by only 3%. For the leading seven emitters, 2030 estimates have slightly decreased, flatlined or increased (see 'The seven top emitters').

No single model can predict the future, and such analyses by necessity exclude the most recent developments. Nevertheless, it is clear that, collectively, current policies will not limit global warming to well below 2 °C, let alone 1.5 °C, as agreed in Paris.

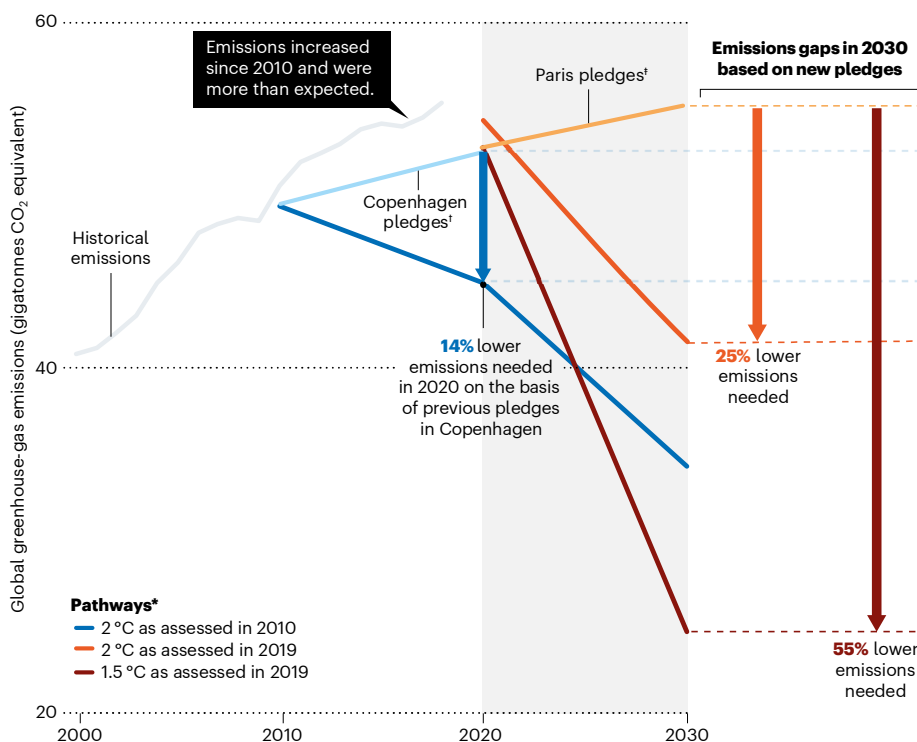
Clearly, the annual audit of the emissions gap has not altered poor performance. The gap concept has nonetheless proved useful. The reports and numbers have continuously informed the UN climate summits¹¹ and the emissions gap was noted as a serious concern when parties were adopting the Paris agreement⁹.

Transformative action

Fundamental policy transformations have begun to appear in some sectors, countries, regions, cities and businesses over the past ten years. These innovations seek to achieve the UN's Sustainable Development Goals (SDGs), including climate ones. Slashing emissions now requires 'leaving no one behind'.

MORE AND FASTER

Insufficient climate action during the past decade means that transformational development pathways are now required to reduce greenhouse-gas emissions on time.



*Median of scenarios that meet the temperature goals at global least costs.

⁹Projected emissions in 2020 based on parties' pledges made in 2009–10 for the Copenhagen accord.

¹⁰Projected emissions in 2030 based on pledges made in 2015 for the Paris agreement. Most parties have not updated their pledges, but are expected to do so in 2020.

Emissions are aggregated using global-warming potentials from the 2007 IPCC Fourth Assessment Report. Values from the UNEP Emissions Gap Report 2010 were converted using a correction factor based on the difference in 2010 global CO₂-equivalent emissions.

To recognize, monitor and understand these advances, the gap reports have included examples. Some are discussed here.

Ambitious action. Most encouragingly, a wealth of agile nations, regions, cities and businesses have promised or made radical changes since the Paris agreement (see 'Action gap' and Supplementary Information. See also go.nature.com/2t22tth). At the last count, net-zero

“Current policies will not limit global warming to well below 2 °C, let alone 1.5 °C, as agreed in Paris.”

emissions goals have been set or are being considered by 76 countries or regions (the European Union is the largest) and 14 sub-national regions or states (the largest being California); some locations have begun implementation. Together, these places account for about 21% of global greenhouse-gas emissions^{12,13}.

Fifty-three countries and 31 states and regions have explicitly committed to an emissions-free electricity sector. Seven additional countries have done so implicitly by aiming for net-zero greenhouse-gas emissions. Together, these account for around 18% of global electricity generation¹⁴. Twenty-one countries, 5 regions and more than 52 cities have

committed to make all vehicles emissions-free. Individual examples also exist for sectors in which reaching zero emissions was thought to be difficult, such as heavy industry and aviation. Steel giants ThyssenKrupp in Essen, Germany, and SSAB in Stockholm are aiming for zero-emissions steel production by 2050 and 2045, respectively. The building-materials company Heidelberg Cement, headquartered in Germany, is aiming for zero-emissions cement production by 2050. For aviation, Norway and Scotland hope to make short-haul and domestic flights zero emissions by 2040.

Renewables. Costs of renewable energy are falling faster than expected¹⁵. Renewables are currently the cheapest source of new power generation in most of the world. Solar and wind power will be financially more competitive than will existing coal plants by next year¹⁵. These cost declines, and those of battery storage, are opening up possibilities for large-scale, low-carbon electrification.

Coal consumption. The rise of renewable energy can – must – facilitate a move away from coal. Emerging economies that depend on coal, such as China and India, have begun to address consumption by adjusting the fuel's price, capping its consumption, reducing plans for new coal-fired power plants and supporting renewables. Much more must be done, and quickly – while addressing poverty, energy access and urbanization^{16–18}.

Comment

UNSDGs. Actions to reduce greenhouse-gas emissions are essential for achieving food security, healthy lives and many other SDGs, as confirmed by a growing body of research^{10,19}. For example, renewable energy cuts air pollution, and improves health and energy security compared with fossil fuels.

Closing the gap

These few success stories must be scaled up and mirrored with progress in every sector. The fact that reductions in greenhouse-gas emissions are a prerequisite to achieving sustainable development must propel action.

The gap is so huge that governments, the private sector and communities need to switch into crisis mode, make their climate pledges more ambitious and focus on early and aggressive action. Otherwise, the Paris agreement's long-term goals are out of reach. We do not have another ten years.

The authors

Niklas Höhne is partner at NewClimate Institute, Cologne, Germany, professor for mitigation of greenhouse gases at the

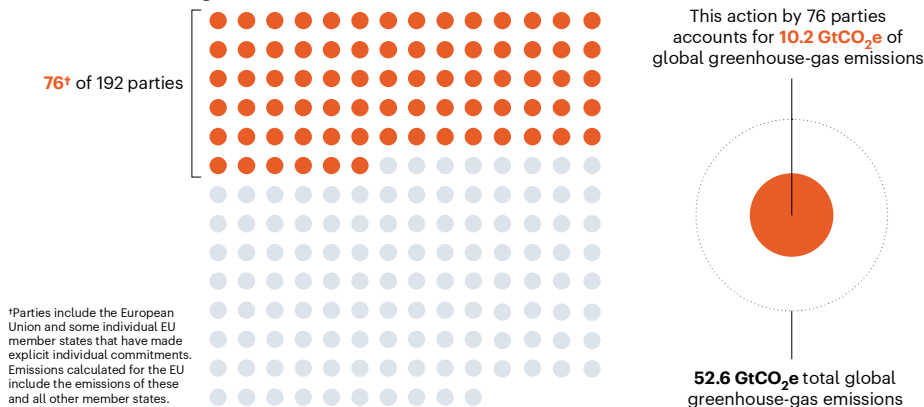
University of Wageningen, the Netherlands, and affiliate at Lawrence Berkeley National Laboratory, California, USA. **Michel den Elzen** is a senior researcher at the PBL Netherlands Environmental Assessment Agency, The Hague, the Netherlands. **Joeri Rogelj** is lecturer in climate change at the Grantham Institute, Imperial College London, UK, and senior research scholar at the International Institute for Applied Systems Analysis, Laxenburg, Austria. **Bert Metz** is former co-chair of the IPCC Working Group on Mitigation and fellow at the European Climate Foundation, The Hague, the Netherlands. **Taryn Fransen** is senior fellow at the World Resources Institute, Washington DC, USA. **Takeshi Kuramochi** is senior climate-policy researcher at NewClimate Institute, Cologne, Germany, and affiliated with the Copernicus Institute of Sustainable Development, Utrecht University, the Netherlands. **Anne Olhoff** is strategic director of climate planning and policy at UNEP DTU Partnership, Technical University of Denmark, Copenhagen, Denmark. **Joseph Alcamo** is professor of environmental systems science and director of the Sussex Sustainability Research Programme at the University of Sussex,

Brighton, UK. **Harald Winkler** is professor in energy and climate policy at the University of Cape Town, South Africa. **Sha Fu** is an associate professor in climate economics at Energy Foundation China, Beijing, China. **Michiel Schaeffer** is science director at Climate Analytics, Berlin, Germany, and affiliated with Environmental Systems Analysis Group, Wageningen University, the Netherlands. **Roberto Schaeffer** is professor of energy economics at the Federal University of Rio de Janeiro, Brazil. **Glen P. Peters** is research director at the CICERO Center for International Climate Research, Oslo, Norway. **Simon Maxwell** is senior research associate at the Overseas Development Institute, London, UK. **Navroz K. Dubash** is professor and coordinator of the Initiative on Climate, Energy and Environment at the Centre for Policy Research, New Delhi, India. e-mail: n.hoehne@newclimate.org

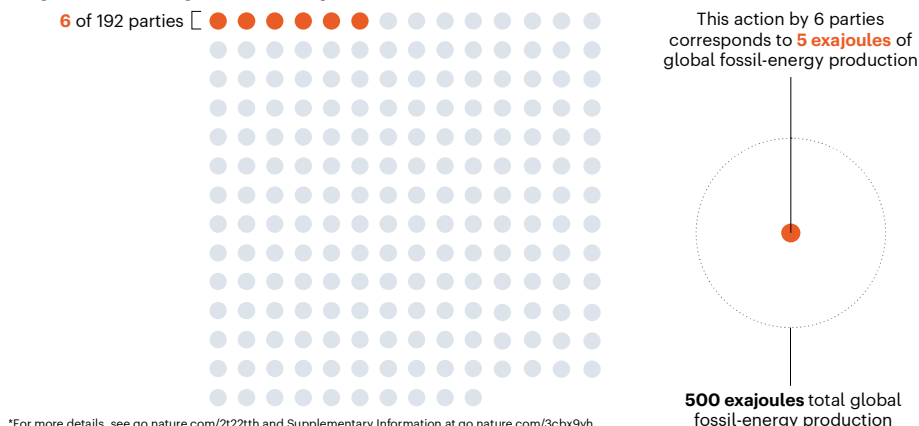
ACTION GAP

Although 192 parties pledged various emissions targets under the Paris climate agreement, commitments to specific actions remain sparse*.

Net-zero emissions goals set



Stop fossil-fuel exploration and production



*For more details, see go.nature.com/2t22tth and Supplementary Information at go.nature.com/3cbx9yh.

1. United Nations Environment Programme. *The Emissions Gap Report. A Preliminary Assessment* (UNEP, 2010).
2. United Nations Environment Programme. *Emissions Gap Report 2019* (UNEP, 2019).
3. Rogelj, J. et al. *Nature* **534**, 631–639 (2016).
4. Höhne, N. et al. *Clim. Pol.* **12**, 356–377 (2011).
5. Rogelj, J. et al. *Nature* **464**, 1126–1128 (2010).
6. Olivier, J. G. J. & Peters, J. A. H. W. *Trends in Global CO₂ and Total Greenhouse Gas Emissions* (PBL Netherlands Environmental Assessment Agency, 2019).
7. United Nations Framework Convention on Climate Change. *Report of the Conference of the Parties on its Fifteenth Session [the Copenhagen accord 2009]* (UNFCCC, 2010).
8. United Nations Framework Convention on Climate Change. *Report of the Conference of the Parties on its Sixteenth Session [the Cancun agreement 2010]* (UNFCCC, 2011).
9. United Nations Framework Convention on Climate Change. *Report of the Conference of the Parties on its Twenty-first Session [the Paris agreement 2015]* (UNFCCC, 2016).
10. Intergovernmental Panel on Climate Change. *Global Warming of 1.5 °C* (IPCC, 2018).
11. Buhr, K., Roth, S. & Stigson, P. *Sustainability* **6**, 794–811 (2014).
12. Olivier, J. G. J. & Peters, J. A. H. W. *Trends in Global CO₂ and Total Greenhouse Gas Emissions* (PBL Netherlands Environmental Assessment Agency, 2018).
13. FAOSTAT. Land Use Emissions. Available at <http://www.fao.org/faostat/en/#data/GL> (2019).
14. International Energy Agency. *World Energy Balances* (IEA, 2019).
15. International Renewable Energy Agency. *Renewable Power Generation Costs in 2018* (IRENA, 2019).
16. Jewell, J., Vinichenko, V., Nacke, L. & Cherp, A. *Nature Clim. Change* **9**, 592–597 (2019).
17. Smouse, S. M., Jones, A., Fapohunda, B. O., Render, M. & Hindman, J. W. in *Proc. ASME 2018 Power Conference*, V001T05A003 (Am. Soc. Mech. Eng., 2018).
18. Arnt, D., Arndt, C., Miller, M., Tarp, F. & Zinaman, O. (eds) *The Political Economy of Clean Energy Transitions* (Oxford Univ. Press, 2017).
19. Antwi-Agyei, P., Dougill, A. J., Agyekum, T. P. & Stringer, L. C. *Clim. Pol.* **18**, 1296–1312 (2018).
20. United Nations Environment Programme. *Emissions Gap Report 2015* (UNEP, 2015).
21. International Energy Agency. *World Energy Outlook 2015* (IEA, 2015).
22. International Energy Agency. *World Energy Outlook 2018* (IEA, 2018).
23. Kuramochi, T. et al. *Greenhouse Gas Mitigation Scenarios for Major Emitting Countries* (NewClimate Institute, 2018).
24. Rochedo, P. R. R. et al. *Nature Clim. Change* **8**, 695–699 (2018).
25. den Elzen, M. et al. *Energy Pol.* **126**, 238–250 (2019).

Supplementary Information accompanies this article: see go.nature.com/3cbx9yh.

SOURCES: 2017 EMISSIONS: REFS 12,13; 2017 ENERGY PRODUCTION: REF. 14.

Correspondence

Fast peer review for COVID-19 preprints

The public call for rapid sharing of research data relevant to the COVID-19 outbreak (see go.nature.com/2t1lyp6) is driving an unprecedented surge in (unrefereed) preprints. To help pinpoint the most important research, we have launched Outbreak Science Rapid PREreview, with support from the London-based charity Wellcome. This is an open-source platform for rapid review of preprints related to emerging outbreaks (see <https://outbreaksci.prereview.org>).

These reviews comprise responses to short, yes-or-no questions, with optional commenting. The questions are designed to capture structured, high-level input on the importance and quality of the research, which can be aggregated across several reviews. Scientists who have ORCID IDs can submit their reviews as they read the preprints (currently limited to the medRxiv, bioRxiv and arXiv repositories). The reviews are open and can be submitted anonymously.

Outbreaks of pathogens such as the SARS-CoV-2 coronavirus that is responsible for COVID-19 move fast and can affect anyone. Research to support outbreak response needs to be fast and open, too, as do mechanisms to review outbreak-related research. Help other scientists, as well as the media, journals and public-health officials, to find the most important COVID-19 preprints now.

Michael A. Johansson Outbreak Science, San Juan, Puerto Rico.
michael@outbreakscience.org

Daniela Saderi PREreview, Portland, Oregon, USA.

Ethics of editing human genomes

As leaders of the national ethics committees of France and Germany, and of the UK Nuffield Council on Bioethics, we consider that the moral and societal issues raised by developments in heritable human-genome editing demand a level of public ethical reflection that current initiatives fail to meet.

In a joint statement, we call on governments and stakeholders worldwide to ensure that heritable genome editing is brought within the control of relevant public authorities (see go.nature.com/3ck1mc). Furthermore, no clinical applications should be considered until there has been broad societal debate about their acceptability and until research has reduced the considerable risks of clinical use to an acceptable level. Measures must be in place to ensure that these risks can be properly assessed and monitored.

Moreover, any ethically permissible application of human genome editing should not increase disadvantage, discrimination or division in society. The large range of conceivable applications, as well as their implications for families, society and future generations, calls for cautious, responsible and transparent governance (see also go.nature.com/3c9fe1).

David Archard Nuffield Council on Bioethics, London, UK.
d.archard@qub.ac.uk

Peter Dabrock German Ethics Council, Berlin, Germany.

Jean-François Delfraissy National Consultative Ethics Committee for Health and Life Sciences, Paris, France.

Peru's research: CONCYTEC responds

As president of Peru's National Council of Science, Technology and Technological Innovation (CONCYTEC), I disagree that the government is not showing sufficient interest in the country's research (see *Nature* **576**, S65–S67; 2019).

The government's expenditure on research and development has increased over the past decade, and this year sees its highest budget ever, at 214 million soles (US\$63 million; see go.nature.com/2ufux1k, in Spanish). And some public universities are investing their royalties from natural resources such as mining into research infrastructure and projects.

A 2018 report by Elsevier commissioned by CONCYTEC indicates that Peru's field-weighted citation impact in 2013–17 was above the world's average. And, according to SCImago rankings, Peru's research is becoming less dependent on international collaborations, with more than 40% of its publications in 2018 exclusively authored by Peruvian scientists.

Last May, the government passed a law to attract and retain more highly qualified scientists. CONCYTEC, with the support of a World Bank project, incorporated 181 local and foreign researchers into Peruvian institutions in 2019.

Notwithstanding these efforts, we recognize that we still have a long way to go in improving Peru's research.

Fabiola María León-Velarde Servetto CONCYTEC, Lima, Peru.
fleon-velarde@concytec.gob.pe

Social priming: a dubious term

The great replicability mystery of 'social priming' in psychology (*Nature* **576**, 200–202; 2019) turns out to reflect a mundane fact: priming studies (social or non-social) that use reliable methods are highly replicable, whereas those that don't are not. In our view, it is time to dispense with the term once and for all.

Social priming occurs when exposure to a social concept or stimulus affects later behaviour. One problem is that there is no clear social component to much of what is defined as social priming (in priming with numbers or the idea of death, for example). And many studies that are obviously social (such as priming with stereotypes) are excluded.

Furthermore, those studies identified as social priming almost exclusively collect a single response to a single prime per subject, whereas others that collect hundreds of responses to multiple primes are excluded from analyses of social priming. Thus, social-priming studies have less power to detect real effects and are more prone to false positives.

Dozens of priming effects using social stimuli are designed to observe multiple behaviours and are highly replicable. But when a non-social priming study measures only a single response per subject, the effects are – unsurprisingly – weak and unreliable (see A. M. Rivers and J. W. Sherman Preprint at PsyArXiv <http://doi.org/dng4>; 2018).

Jeffrey W. Sherman University of California, Davis, California, USA.
jsherman@ucdavis.edu

Andrew M. Rivers University of British Columbia, Vancouver, Canada.

News & views

Biomechanics

Ahead of the curve in the evolution of human feet

Glen A. Lichtwark & Luke A. Kelly

The longitudinal arch has long been considered a crucial structure that provides stiffness to the human foot. Now the transverse arch is stepping into the spotlight, with a proposed central role in the evolution of human foot stiffness. **See p.97**

Humans evolved to walk and run effectively on the ground using two feet. Our arched foot, which is not a characteristic of other primates, is a unique feature crucial for human bipedalism. The arch provides the foot with the stiffness necessary to act as a lever that transmits the forces generated by leg muscles as they push against the ground. The arch also retains sufficient flexibility to function like a spring to store and then release mechanical energy. On page 97, Venkadesan *et al.*¹ present a new view of how foot stiffness is regulated. Their finding not only has exciting implications for understanding foot evolution, but also provides a possible framework when considering foot health and how to design better footwear.

The foot's longitudinal arch (the arch that runs from the heel to the ball of the foot; Fig. 1) is often credited^{2,3} with the leading role in foot stiffening. The ligaments spanning this arch, including the plantar fascia (or plantar aponeurosis), act like a bowstring to resist arch collapse when force is applied. Moreover, the spring-like mechanical properties of these ligaments contribute substantially to the foot's ability to store and return energy⁴.

However, Venkadesan and colleagues present the idea that another arch component, the transverse arch (the part of the arch that curves across the foot at the base of the metatarsal bones; Fig. 1) is at least as important for foot stiffness as is the longitudinal arch, if not more so. The authors provide evidence for how transverse-arch curvature might help prevent foot bending and therefore increase foot stiffness. An analogy for this proposed stiffening mechanism is the way that a pizza slice becomes less floppy if the slice's outer crust is curled up.

Venkadesan *et al.* initially took a theoretical approach to investigate the role of transverse

curvature in stiffening the foot. Modelling an elastic shell, the authors demonstrated that, if the transverse curvature of the shell increased, this increased the stiffness of the shell in the longitudinal direction. Venkadesan and colleagues derived a parameter for curvature and longitudinal stiffness (independent of other factors such as shell size and thickness), and show that a distinct transition point exists beyond which the amount of curvature directly influences the longitudinal stiffness. A similar relationship exists for a physical model consisting of discrete rigid elements (analogous to the metatarsal bones) connected by springs (corresponding to ligaments).

To test whether this model might be relevant

to the stiffness of the human foot arch, the authors examined human cadaver specimens (frozen after death and then thawed to combat stiffening due to rigor mortis) and cut ligaments in the transverse arch that are expected to be crucial for coupling the curvature of this arch to foot stiffness. Venkadesan *et al.* then assessed the foot's vertical deformation when loads were applied. Cutting the transverse ligaments reduced foot stiffness by the remarkable value of more than 40%. By comparison, previous research⁴ indicates that cutting the foot's plantar fascia, which spans the longitudinal arch, reduces stiffness by just 23%. Venkadesan and colleagues' data therefore suggest that transverse ligaments make a substantial contribution to overall foot stiffness. When bearing a load, the foot's transverse ligaments are presumably stretched by the resultant spreading out of the metatarsals at the ball of the foot. The authors suggest that this ligament stretching is a direct result of transverse-arch curvature.

Venkadesan and co-workers examined the evolution of the transverse arch across different primates, including various species of extinct hominin (those species more closely related to humans than to chimpanzees). As in other work⁵ investigating foot evolution, Venkadesan *et al.* focused on the amount of torsion (twist) in the fourth metatarsal bone. They estimated the curvature of the transverse arch and determined which species would probably have had sufficient

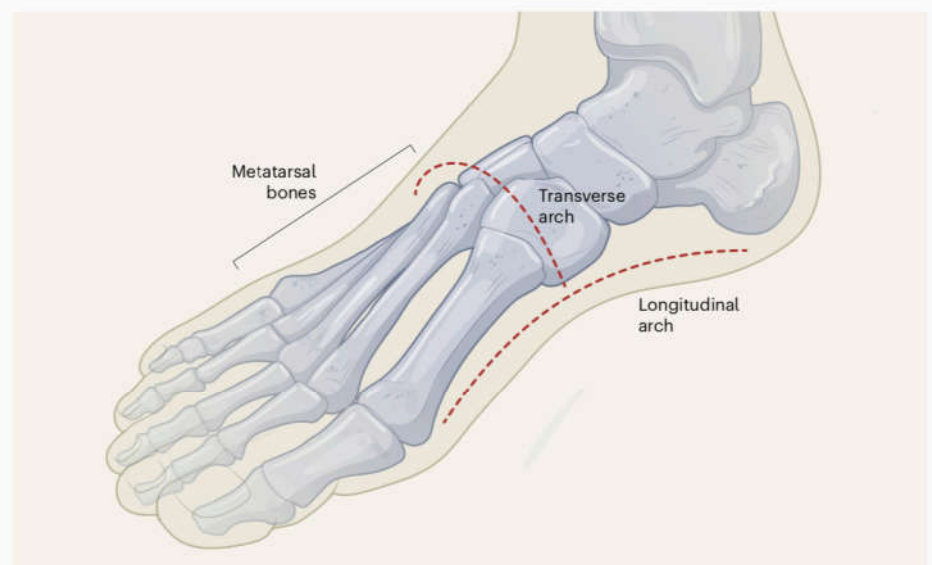


Figure 1 | Human foot arches. The longitudinal arch of the human foot has been proposed^{2,3} to have a key role in providing stiffness for the foot, an attribute that enables humans to walk on the ground on two feet. Venkadesan *et al.*¹ report that another foot arch — the transverse arch, which is in the vicinity of the metatarsal bones — makes a major contribution to foot stiffness.

curvature to induce stiffening of this arch to an extent similar to that of modern humans. For example, the authors examined the species *Australopithecus afarensis*. This species existed more than three million years ago, and whether it walked upright in a human-like fashion is debated^{6–8}. Venkadesan *et al.* report that the transverse arch of *A. afarensis* was less curved than that of a human foot and thus, according to their model, probably less stiff. However, the authors correctly emphasize that such curvature alone cannot be used reliably to infer movement capabilities, and other mechanisms might stiffen the foot sufficiently to allow a human-like gait.

The curvature of transverse arches in human populations probably spans a wide range of values. Some people have noticeably flat feet whereas others have a high arch. Perhaps those with flat feet have less curvature of their transverse arch and thus potentially reduced stiffness in their feet compared with those whose feet are less flat. But it is also possible that people with flat feet have sufficient transverse-arch curvature to compensate for their low longitudinal arch, thereby maintaining sufficient stiffness for effective walking and running. Given that Venkadesan and colleagues' work did not directly test whether there is a relationship between transverse-arch curvature and the stiffness of the human foot, it remains to be determined whether the range of differences in human transverse-arch curvature is a crucial functional parameter to explain foot stiffness.

The range of curvature of the arch of human feet suggested by Venkadesan *et al.* would indicate that a nearly twofold change in stiffness is possible as a result of natural variation in curvature of the transverse arch from one person to the next. However, any relationship between transverse-arch curvature and stiffness is probably not enough to completely explain the regulation of foot stiffness, and other factors will also need to be considered – for example, the stiffness of the plantar fascia or the potential for muscles to actively regulate arch stiffness. As such, caution is necessary before relying on this curvature parameter alone as the key variable in assessing human foot stiffness.

The fields of evolutionary biology, sports science and medicine have largely neglected the transverse arch when trying to explain the managements of loads applied to the foot. Venkadesan and colleagues' research suggests a new mechanism that links foot form and function and sets the scene for a possible shift in how the human foot is considered. More research will be needed to better understand how the transverse arch contributes to human locomotor performance, including determining what its contribution is to an individual's foot stiffness and whether this provides any mechanical or energetic benefits.

It is conceivable that new treatments that take advantage of transverse-arch curvature to modulate foot stiffness could be developed for various foot disorders. Perhaps even more exciting are the implications of this work for efforts to mimic a human foot when designing prosthetic limbs or legged robots.

Glen A. Lichtwark and **Luke A. Kelly** are at the University of Queensland, School of Human Movement and Nutrition Sciences, St Lucia, Queensland 4072, Australia.
e-mails: g.lichtwark@uq.edu.au;
l.kelly3@uq.edu.au

1. Venkadesan, M. *et al. Nature* **579**, 97–100 (2020).
2. Pontzer, H. *Curr. Biol.* **27**, R613–R621 (2017).
3. Holowka, N. B. & Lieberman, D. E. *J. Exp. Biol.* **221**, eb174425 (2018).
4. Ker, R. F., Bennett, M. B., Bibby, S. R., Kester, R. C. & Alexander, R. M. *Nature* **325**, 147–149 (1987).
5. Ward, C. V., Kimbel, W. H. & Johanson, D. C. *Science* **331**, 750–753 (2011).
6. Jungers, W. L. *Nature* **297**, 676–678 (1982).
7. Ward, C. V. *Am. J. Phys. Anthropol.* **119** (Suppl. 35), 185–215 (2002).
8. Hatala, K. G., Demes, B. & Richmond, B. G. *Proc. R. Soc. B* **283**, 20160235 (2016).

The authors declare competing financial interests: see go.nature.com/2vefydy for details.

This article was published online on 26 February 2020.

Artificial intelligence

In-sensor computing for machine vision

Yang Chai

An image-sensor array has been developed that acts as its own artificial neural network to capture and identify optical images simultaneously, processing the information rapidly without needing to convert it to a digital format. **See p.62**

Sight is one of our most vital senses. Biologically inspired machine vision has developed rapidly in the past decade, to the point that artificial systems can 'see' in the sense of gaining valuable information from images and videos^{1,2}, although human vision remains much more efficient. On page 62, Menzel *et al.*³ report a design for a visual system that, rather like the brain, can be trained to classify simple images in nanoseconds.

Modern image sensors such as those in digital cameras are based on semiconductor (solid-state) technology and were developed in the early 1970s; they fall into two main types, known as charge-coupled devices and active-pixel sensors⁴. These sensors can faithfully capture visual information from the environment, but generate a lot of redundant data. This vast amount of optical information is usually converted to a digital electronic format and passed to a computing unit for image processing.

The resulting movement of massive amounts of data between sensor and processing unit results in delays (latency) and high power consumption. As imaging rates and numbers of pixels grow, bandwidth limitations make it difficult to send everything back to a centralized or cloud-based computer rapidly enough for real-time processing and decision-making – which is especially important for delay-sensitive applications such as driverless vehicles,

robotics or industrial manufacturing.

A better solution would be to shift some of the computational tasks to the sensory devices at the outer edges of the computer system, reducing unnecessary data movement. And because sensors normally produce analog (continuously varying) outputs, analog processing would be preferable to digital: analog-to-digital conversion is notoriously time- and energy-consuming.

To mimic the brain's efficient processing of information, biologically inspired neuromorphic engineering adopts a computing architecture that has highly interconnected elements (neurons, connected by synapses), allowing parallel computing (Fig. 1a). These artificial neural networks can learn from their surroundings by iteration – for instance, learning to classify something after being shown known examples (supervised learning), or to recognize a characteristic structure of an object from input data without extra information (unsupervised learning). During learning, an algorithm repeatedly makes predictions and strengthens or weakens each synapse in the network until it reaches an optimum setting.

Menzel and co-workers implement an artificial neural network directly in their image sensor. On a chip, they construct a network of photodiodes – tiny, light-sensitive units, each consisting of a few atomic layers of tungsten diselenide. This semiconductor's response

to light can be increased or decreased by altering an applied voltage, so that the sensitivity of each diode can be individually tuned. In effect, this turns the photosensor network into a neural network (Fig. 1b) and allows it to carry out simple computational tasks. Changing the light responsiveness of a photodiode alters the connection strength – the synaptic weight – in the network. Thus, the device combines optical sensing with neuromorphic computing.

The authors arrange the photodiodes into a square array of nine pixels, with three diodes to each pixel. When an image is projected on to the chip, various diode currents are produced, combined and read. The hardware array provides a form of analog computing: each photodiode generates an output current that is proportional to the incident light intensity, and the resulting currents are summed along a row or column, according to Kirchhoff's law (a fundamental rule of currents in circuits).

The array is then trained to perform a task. The discrepancy between the currents produced by the array and the predicted currents (the currents that would be produced if the array responds correctly to the image, for a given task) is analysed off-chip and used to adjust the synaptic weight for the next training cycle. This learning stage takes up time and computing resources, but, once trained, the chip performs its set task rapidly.

Using different algorithms for the neural network, the authors demonstrate two neuromorphic functions. The first is classification: their 3×3 array of pixels can sort an image into one of three classes that correspond to three simplified letters, and thus identify which letter it is in nanoseconds. This relatively simple task is just a proof of concept, and could be extended to recognizing more-complicated images if the array size were scaled up.

The second function is autoencoding: the computing-in-sensor array can produce a simplified representation of a processed image by learning its key features, even in the presence of signal noise. The encoded version contains only the most essential information, but can be decoded to reconstruct an image close to the original.

There is more to be done before this promising technology can be used in practical applications. A neuromorphic visual system for autonomous vehicles and robotics will need to capture dynamic images and videos in three dimensions and with a wide field of view. Currently used image-capture technology usually translates the 3D real world into 2D information, thereby losing movement information and depth. The planar shape of existing image-sensor arrays also restricts the development of wide-field cameras⁵.

Imaging under dim light would be difficult for the device described by the authors. A redesign would be needed to improve light

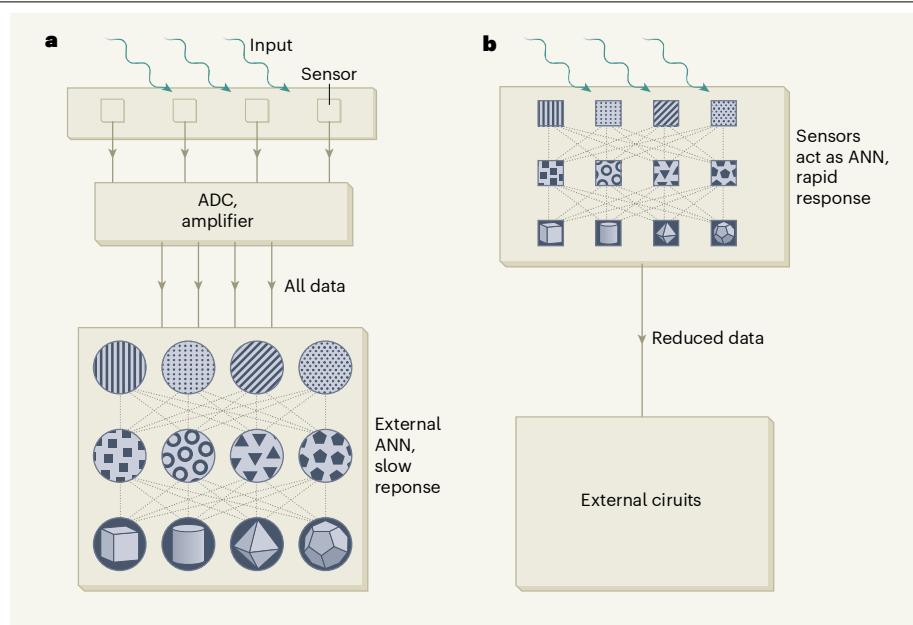


Figure 1 | Computing within a vision sensor for intelligent and efficient preprocessing. **a**, In conventional artificial-intelligence (AI) vision sensors, signals are collected from light-responsive sensors, converted from analog to digital form (ADC, analog-to-digital converter), amplified and then fed as inputs to an external artificial neural network (ANN) – layers of interconnected computational units (circles) whose connections can be adjusted, allowing the network to be trained to perform tasks such as classifying images. An input layer of the ANN receives signals encoding simple physical elements (represented here by dots and lines); in subsequent layers, these are optimized to mid-level features (simple shapes); and refined images are formed at the output layer (3D shapes). The overall response can be slow and energy-hungry. **b**, Mennel *et al.*³ report a system in which interconnected sensors (squares) on a chip not only collect signals, but also work as an ANN to recognize simple features, reducing movement of redundant data between sensors and external circuits.

absorption in the thin semiconductor and to increase the range of light intensities that can be detected. Furthermore, the reported design requires high voltages and consumes a lot of power; by comparison, the energy consumption per operation in a biological neural network is at the sub-femtojoule level (10^{-15} to 10^{-13} joules)⁶. It would also be useful to expand the response to ultraviolet and infrared light, to capture information unavailable in the visible spectrum⁷.

The thin semiconductors used are difficult to produce uniformly over large areas, and are hard to process so that they can be integrated with silicon electronics, such as external circuits used for readout or feedback control. The speed and energy efficiency of devices that use these sensors will be dominated not by the image-capturing process, but by data movement between sensors and external circuits. Moreover, although the computing-in-sensor unit collects and computes data in the analog domain, reducing analog-to-digital conversions, the peripheral circuits still suffer from other intrinsic delays. The sensors and external circuits will need to be co-developed to decrease the latency of the entire system.

Mennel and colleagues' computing-in-sensor system should inspire further research into artificial-intelligence (AI) hardware. A few companies have developed AI

vision chips based on silicon electronics⁸, but the chips' intrinsic digital architecture leads to problems of latency and power efficiency.

More broadly, the authors' strategy is not limited to visual systems. It could be extended to other physical inputs for auditory, tactile, thermal or olfactory sensing^{9–11}. Development of such intelligent systems, together with the arrival of the 5G fast wireless network, should allow real-time edge (low-latency) computing in the future.

Yang Chai is in the Department of Applied Physics, Hong Kong Polytechnic University, Kowloon, Hong Kong, P. R. China. e-mail: ychai@polyu.edu.hk

1. Mead, C. *Proc. IEEE* **78**, 1629–1636 (1990).
2. Kyuma, K. *et al. Nature* **372**, 197–198 (1994).
3. Mennel, L. *et al. Nature* **579**, 62–66 (2020).
4. Bigas, M., Cabruja, E., Forest, J. & Salvi, J. *Microelectr. J.* **37**, 433–451 (2006).
5. Choi, C. *et al. Nature Commun.* **8**, 1664 (2017).
6. Laughlin, S. B., de Ruyter van Steveninck, R. R. & Anderson, J. C. *Nature Neurosci.* **1**, 36–41 (1998).
7. Zhang, K. *et al. ACS Nano* **10**, 3852–3858 (2016).
8. Davies, M. *et al. IEEE Micro* **38**, 82–99 (2018).
9. Wan, C. *et al. Adv. Mater.* **30**, 1801291 (2018).
10. Kim, Y. *et al. Science* **360**, 998–1003 (2018).
11. Qu, T. Y. *et al. Adv. Mater.* **32**, 1907288 (2020).

Microbiology

One gene to rule them all in a chronic brain infection

Eva-Maria Frickel

A gene has been found that controls the conversion of the parasite *Toxoplasma gondii* into a form that chronically infects the human brain. The discovery could aid the design of therapies to eliminate this currently untreatable infection.

It is estimated that around one-third of the global human population¹ is infected with the single-celled organism *Toxoplasma gondii*, a parasite that can be ingested in food or picked up from activities such as gardening². The parasite needs to differentiate into a chronic-stage form to establish a permanent infection in brain and muscle tissue, but how this parasite conversion occurs has been a mystery. Writing in *Cell*, Waldman *et al.*³ report the identification of a gene that encodes a master regulator of this differentiation event.

Toxoplasma gondii can infect any warm-blooded animal. Human infection can occur through eating undercooked meat from infected livestock or by ingesting contaminated food or water. Within a couple of weeks of entering its host, *T. gondii* is converted from a form called an acute-stage tachyzoite into a bradyzoite, which establishes a chronic infection (Fig. 1). A bradyzoite forms a cyst that resides in host cells and is surrounded by a thick wall of proteins and sugars. The wall is a formidable barrier that makes the cyst inaccessible and thwarts its elimination by drugs or the host's immune system.

Although *T. gondii* infection is widespread in human populations, it is often harmless, being in the relatively quiescent state of bradyzoites that have not reverted to the activated tachyzoite form associated with disease. However, *T. gondii* infection can be life-threatening for unborn fetuses or for people whose immune systems are compromised. Moreover, in the United States, 2% of *T. gondii* infections result in sight problems or blindness owing to ocular damage caused by treatment-resistant parasites⁴.

To uncover the signal that controls the formation of bradyzoites, Waldman and colleagues engineered *T. gondii* to express a green fluorescent protein if such cysts formed. Monitoring the fluorescent protein using microscopy and cell-sorting technologies offered a way of assessing whether the parasite had differentiated into the form associated with chronic infection.

Exposure to stress-inducing treatment in culture conditions, such as an alkaline pH, made the parasite differentiate into bradyzoites. The authors used the gene-editing tool CRISPR to disrupt selected genes to assess

“The parasite needs to differentiate into a chronic-stage form to establish a permanent infection.”

whether any of them affected differentiation. The results were stunning and clear. The disruption of only one targeted gene – which the authors call bradyzoite-formation

deficient 1 (*BFD1*) – prevented the formation of bradyzoites.

BFD1 encodes a transcription-factor protein belonging to a family known as Myb-domain-containing proteins. Waldman *et al.* demonstrate that the Myb domain of *BFD1* protein drives *T. gondii* differentiation. This is particularly intriguing because another Myb-domain-containing protein controls chronic-stage cyst formation in the parasite *Giardia lamblia*⁵. Moreover, a related member of the Myb-domain-containing protein family enables *Plasmodium* parasites to develop in red blood cells⁶. In addition to *BFD1*, *T. gondii* encodes 13 other Myb-domain-containing proteins. Identifying their functions and determining whether any of them aid the infection process should be a priority.

Waldman and co-workers report that *T. gondii* lacking *BFD1* fail to establish a chronic infection in mice. When investigating the regulation of *BFD1* expression, the authors made the counter-intuitive discovery that the messenger RNA that encodes *BFD1* was expressed at a similar level during both the acute and chronic stages of infection. The presence of *BFD1* is sufficient to drive parasite differentiation into a bradyzoite, and the mRNA encoding *BFD1* is preferentially translated into protein during the chronic stage of infection. Leveraging this finding, Waldman *et al.* engineered *T. gondii* to express a form of *BFD1* that is unstable unless a specific compound is also given. Consistent with the authors' model, the compound-mediated stabilization of *BFD1* caused the parasite to form a bradyzoite.

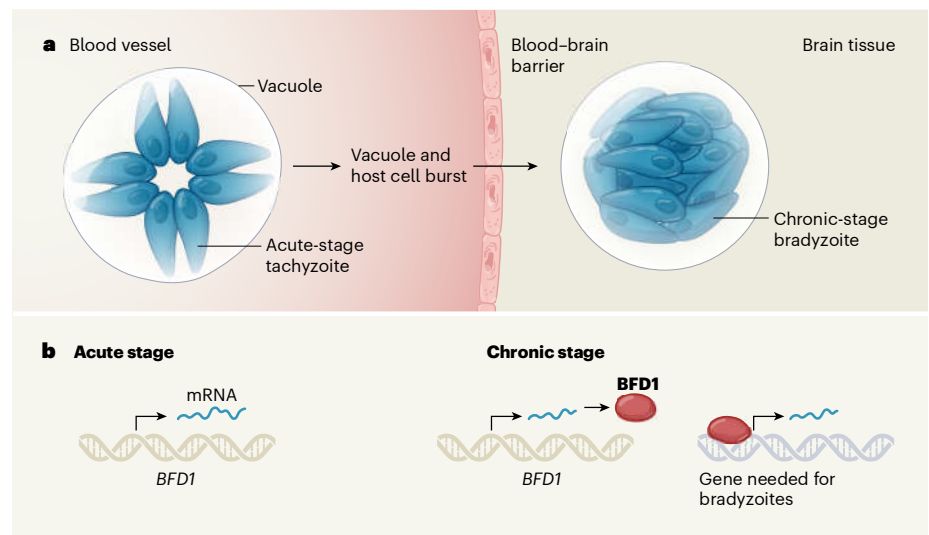


Figure 1 | How *Toxoplasma gondii* parasites differentiate to cause a chronic infection. **a**, *Toxoplasma gondii* infects humans, and can be life-threatening. During the initial stages of infection, the parasite exists in the bloodstream in a form called an acute-stage tachyzoite, which is in a vacuole. It is taken up by a host cell (not shown) and the cell and vacuole subsequently burst. The parasite enters the brain and gives rise to a chronic infection. Such infection occurs when the parasite differentiates into a form called a chronic-stage bradyzoite. **b**, Waldman *et al.*³ report that the gene *BFD1* is required for this differentiation step. In both the acute and chronic stages of infection, this gene is transcribed into messenger RNA. However, the encoded protein *BFD1* is preferentially made during chronic infection. *BFD1* is a transcription-factor protein that can drive the expression of genes needed for the formation of bradyzoites.

This discovery raises the question of how the translation of the mRNA that encodes BFD1 is regulated, possibly in response to stress, to trigger chronic infection.

As expected, the authors observed that *T. gondii* parasites differentiated into bradyzoites after several rounds of replication in host cells *in vitro* under stressful conditions (*in vivo* stress arises, in part, from the host's mounting immune response). This process was not synchronous across all parasites being cultured or even for those in one host cell. The researchers therefore used single-cell RNA profiling of wild-type and BFD1-deficient parasites to assess gene-expression profiles associated with the differentiation event. They also investigated the regions of the parasite genome to which BFD1 binds. Gratifyingly, as expected for a transcription factor, BFD1 bound to gene regions called transcription start sites, and, in particular, to those in a large set of genes that the authors had identified as being expressed at higher than normal levels during differentiation.

Many questions remain unanswered regarding BFD1's regulation of differentiation, and how it might act upstream of a group of previously identified transcription factors called ApiAP2s, which are important, but not sufficient, for differentiation⁷. Considering that BFD1 is probably regulated by translational control, approaches that determine the RNA content of single cells might not be enough to identify the full cohort of factors driving differentiation. Another way to investigate translational control is to profile RNAs bound to the translational machinery of the ribosome complex. This method has already been used for *T. gondii*^{8,9}, and should be enlisted to study bradyzoites.

Bradyzoites can now be maintained in host cells grown *in vitro* without adversely affecting the host cells, opening many vistas for future experiments. Particularly exciting is the possibility of analysing bradyzoites during brain infection by using an approach that harnesses stem-cell technologies, such as those that produce neuronal stem cells. CRISPR provides a way of testing the role of host genes, and this method can also target *T. gondii* both *in vitro* and *in vivo*^{10–12}. The availability of these tools sets the stage for new discoveries about the interplay between the parasite, host and immune system throughout the acute and chronic stages of infection. The development of artificial-intelligence methods that enable computer-driven assessments of complex and subtle differences in images of *T. gondii* offers another way of assessing the infection process¹³.

Given that bradyzoites are the most relevant and challenging stage of the *T. gondii* life cycle to tackle for the treatment of the human disease, targeting BFD1 shows real potential for making progress in the development of drugs

or vaccines. The discovery of one gene that can rule them all moves us closer to solving the riddle of this chronic infection.

Eva-Maria Frickel is at the Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK, and at the Francis Crick Institute, London, UK. e-mail: e.frickel@bham.ac.uk

1. Montoya, J. G. & Liesenfeld, O. *Lancet* **363**, 1965–1976 (2004).
2. Furtado, J. M., Smith, J. R., Belfort, R., Gattey, D. & Winthrop, K. L. *J. Glob. Infect. Dis.* **3**, 281–284 (2011).

3. Waldman, B. S. et al. *Cell* **180**, 359–372 (2020).
4. Jones, J. L. et al. *Clin. Infect. Dis.* **60**, 271–273 (2015).
5. Sun, C.-H., Palm, D., McArthur, A. G., Svård, S. G. & Gillin, F. D. *Mol. Microbiol.* **46**, 971–984 (2002).
6. Bosch, C. et al. *Mol. Biochem. Parasitol.* **138**, 159–163 (2004).
7. Jeffers, V., Tampaki, Z., Kim, K. & Sullivan, W. J. Jr. *Cell. Mol. Life Sci.* **75**, 2355–2373 (2018).
8. Hassan, M. A., Vasquez, J. J., Guo-Liang, C., Meissner, M. & Siegel, T. N. *BMC Genom.* **18**, 961 (2017).
9. Holmes, M. J., Shah, P., Wek, R. C. & Sullivan, W. J. Jr. *mSphere* **4**, e00292-19 (2019).
10. Sidik, S. M., Huet, D. & Lourido, S. *Nature Protocols* **13**, 307–323 (2018).
11. Sangaré, L. O. et al. *Cell Host Microbe* **26**, 478–492 (2019).
12. Young, J. et al. *Nature Commun.* **10**, 3963 (2019).
13. Fisch, D. et al. *eLife* **8**, e40560 (2019).

Structural biology

A self-activating orphan receptor

Brian Krumm & Bryan L. Roth

The first 3D structure of a full-length G-protein-coupled receptor whose natural activator is unknown has been determined, providing insights into an unusual mode of activation and a basis for discovering therapeutics. **See p.152**

G-protein-coupled receptors are the largest class of membrane protein in the human genome, and represent the most abundant pharmaceutical targets. More than 800 such receptors are known in humans, of which perhaps 100 are orphan receptors – those for which the naturally occurring (endogenous) ligand molecules that bind to and activate them have yet to be identified^{1,2}. This lack of understanding of orphan G-protein-coupled receptors (oGPCRs) impedes our ability to exploit their potential as therapeutic targets. On page 152, Lin *et al.*³ close this gap in knowledge by reporting the first 3D structure of a full-length oGPCR, GPR52, in multiple states.

GPR52 is a potential drug target for treating several neuropsychiatric disorders, including Huntington's disease and schizophrenia. When activated, it selectively binds to the G_s family of G proteins inside cells, and thereby stimulates the production of cyclic AMP (cAMP) signalling molecules, which regulate various cellular processes. Efforts to find drugs that target GPR52 would benefit from a greater knowledge of how the receptor couples to G_s and its activation process.

Lin *et al.* began their investigation of the structural basis for GPR52 activation using X-ray crystallography. In their initial studies, the authors used a variety of strategies, including extensive protein engineering, to both stabilize the receptor and enable its production in sufficient quantities to produce high-resolution crystal structures. The

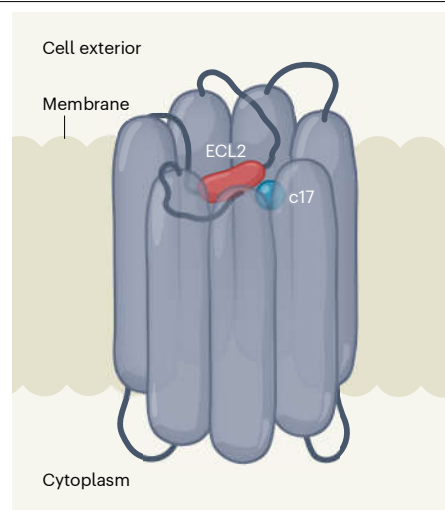


Figure 1 | Binding sites in the receptor GPR52. Lin *et al.*³ report structures of the membrane receptor GPR52, a potential drug target for which the putative naturally occurring agonist – the ligand molecule that activates the receptor – is unknown. The authors find that a region of GPR52 known as extracellular loop 2 (ECL2) binds to a site in the receptor that is analogous to the agonist-binding site in other receptors from the same family. ECL2 seems to activate the receptor, removing the need for an external agonist. The authors also find that the synthetic molecule c17, which activates GPR52, binds to a different region next to the site bound by ECL2, and might therefore be an allosteric modulator (a compound that potentiates the activity of the receptor but does not bind at the agonist-binding site).

researchers thus obtained the structures of human GPR52 in the ligand-free (apo) state and in complex with c17, a synthetic molecule that acts as an agonist (that is, it activates the receptor).

Not unexpectedly, GPR52-apo adopts the GPCR architecture that has been seen in many other structures, involving seven transmembrane domains. Surprisingly, a region of the receptor known as extracellular loop 2 (ECL2) folds into what would normally be the binding site for an endogenous ligand (the orthosteric binding site), where it acts as a lid that blocks the entrance to this site (Fig. 1). Lin *et al.* observed that the activity of GPR52 is significantly diminished when ECL2 is mutated or deleted, indicating that the loop is essential for signalling activity in the receptor's native environment. Meanwhile, the crystal structure of the receptor in complex with c17 suggests that this agonist binds to a 'side pocket' that has not been observed in previously reported structures of GPCRs. The authors therefore speculate that c17 acts allosterically – at a site remote from the orthosteric binding site – to potentiate GPR52's activity.

Remarkably, the authors were then able to form a stable complex of GPR52 with a modified G_s protein in the absence of an agonist, and to obtain the structure of the complex using cryo-electron microscopy. The receptor in this complex has the structural hallmarks of previously visualized, active GPCRs captured in complex with G proteins¹. The arrangement of ECL2 in this active-state structure is the same as in the crystal structure of GPR52-apo, implying that ECL2 acts as a 'tethered agonist' under physiological conditions to facilitate signalling pathways in the absence of an endogenous agonist – similarly to the behaviour of some other GPCRs, such as the PAR1 protease-activated receptor⁴.

Most GPCRs have some basal (constitutive) activity wherein they spontaneously couple to their particular G proteins. The constitutive activity of GPR52 is exceptionally high⁵. Indeed, Lin and colleagues find that GPR52's basal activity is so great that the receptor's ability to signal by increasing cAMP levels is only slightly augmented by the addition of c17.

The authors report that this high level of constitutive activity is achieved by at least two structural features that are unusual for GPCRs: the lack of a binding site for sodium ions, and the occupation of an apparent agonist-binding site by the tethered agonist in ECL2. The sodium-binding site of GPCRs is known to be important for damping constitutive activity⁶, and so the observation that a GPCR that lacks such a site has a high level of basal activity is not entirely surprising. By contrast, the discovery of a tethered agonist that helps to maintain GPR52 in the active state in the absence of

an external agonist is truly striking. The new findings raise the intriguing possibility that, for at least some oGPCRs, the incorporation of agonists within the receptor itself obviates the need for external ligands. Indeed, several other oGPCRs that have high constitutive activities⁵ have been identified, along with others that don't have sodium-binding sites⁶.

It should be kept in mind that – as with all structural studies – Lin and colleagues' work has provided only a few snapshots of the receptor structure. Further biochemical and biophysical studies will be essential to work out the details of GPR52's dynamic behaviour under physiological conditions.

Nevertheless, the authors' high-resolution structures should aid the development of drugs that selectively target GPR52, but avoid other potential drug targets – for instance, by enabling computational studies⁷ in which ultra-large libraries of potential ligands are docked into the binding site revealed by the

structures. Moreover, if the approaches used by Lin *et al.* for the structural elucidation of GPR52 are applied to other oGPCRs that have high constitutive activity^{5,6}, they might transform our understanding of oGPCRs and accelerate their therapeutic exploitation.

Brian Krumm and **Bryan L. Roth** are in the Department of Pharmacology, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27514, USA.
e-mail: bryan_roth@med.unc.edu

1. Wacker, D., Stevens, R. C. & Roth, B. L. *Cell* **170**, 414–427 (2017).
2. Roth, B. L. *Nature Struct. Mol. Biol.* **26**, 535–544 (2019).
3. Lin, X. *et al.* *Nature* **579**, 152–157 (2020).
4. Zhang, C. *et al.* *Nature* **492**, 387–392 (2012).
5. Martin, A. L., Steurer, M. A. & Aronstam, R. S. *PLoS ONE* **10**, e0138463 (2015).
6. Katritch, V. *et al.* *Trends Biochem. Sci.* **39**, 233–244 (2014).
7. Lyu, J. *et al.* *Nature* **566**, 224–229 (2019).

This article was published online on 19 February 2020.

Ecology

Biodiversity theory backed by island bird data

Kostas A. Triantis & Thomas J. Matthews

Analysis of a unique global data set reveals how the species diversity of birds is affected by the properties of archipelagos and offers a way to test an influential theory. Has this improved our understanding of island biodiversity patterns? **See p.92**

The thousands of islands in the Aegean Sea between Greece and Turkey have inspired countless myths and works of literature. This region is also where the word archipelago, which means a group of islands, has its roots. Archipelagos and their constituent islands have long been viewed as natural 'laboratories' for developing and testing theories that aim to answer key questions about biodiversity^{1–5}. On page 92, Valente *et al.*⁶ report an impressive analysis of birds on archipelagos worldwide that provides some of these long-awaited answers.

In the 1960s, the biologists R. H. MacArthur and E. O. Wilson proposed the theory of island biogeography^{7,8}, which is commonly used to explain observed patterns of species richness (the number of different species) on islands. This development marked the dawning of a renaissance for biogeography (the study of species distributions over space and time) that advanced this field from a largely descriptive endeavour to a quantitative and predictive science^{1–5}.

The theory of island biogeography was

inspired by two well-established patterns of species diversity. One pattern is that species richness increases if a greater area is sampled. The other pattern is that the species richness of an island is lower the greater the isolation of the island – the farther away the island is from a potential source of species, such as the closest mainland. The theory of island biogeography predicts that the species richness observed on an island is the result of the interplay between three fundamental processes – extinction, colonization (the dispersal and establishment of species from the continental landmass to an island) and speciation (the generation of new species) – and that these processes depend on island area and isolation. This theory has had a wide-reaching influence on researchers in fields including ecology and conservation biology, and has underpinned the emergence of subdisciplines in these fields, such as macroecology and metapopulation biology^{1–5}.

Yet despite a multitude of studies^{3,5} testing the theory of island biogeography, few have sought to use molecular phylogenies to

directly test on a global scale the dependency of extinction, colonization and speciation on island area and isolation. Valente and colleagues provide such a test. They focused on terrestrial birds, excluding migratory species, and gathered an impressive data set of 491 species across 41 archipelagos worldwide.

Building on their previous work investigating mechanisms that generate island biodiversity⁹, the authors applied an innovative modelling approach that combined molecular phylogenetic data with information on the spatial distribution of birds. The authors obtained genetic data from 90 species across different archipelagos, including 110 island populations not previously sampled. Valente and colleagues also sampled genetic data for the closest mainland-dwelling relatives of several of these island species. After combining their data with pre-existing data, the authors built phylogenetic trees showing the evolutionary relationships between species. Using these phylogenies, they were able to estimate colonization, extinction and speciation rates. The authors also included species known to have been driven to extinction by humans, because excluding such species impedes our understanding of natural processes and biodiversity patterns^{9,10}.

The authors' models, which used rates estimated at the archipelago level, have high explanatory power and confirm several key predictions of the theory of island biogeography – namely, that extinction rates decline with increasing island area, colonization rates decline with increasing distance from the island to the continent, and speciation rates increase with the area and isolation of islands. The authors studied two types of speciation (Fig. 1) separately: anagenesis (in which a new species arises when an island population diverges from its ancestral species on the continent to become a different species³) and cladogenesis (in which an ancestral species splits into two or more different species³). They found that anagenesis increases with island isolation, and cladogenesis increases on larger, more isolated islands. These findings will help future studies that attempt to answer long-debated questions, such as why only certain animal and plant groups speciate extensively, and whether there are upper limits to the species richness and speciation rates in specific regions of the globe³.

Valente and colleagues have not only advanced our understanding of the laws governing species richness on islands, they have also confirmed several predictions of the theory of island biogeography. As the authors mention, the next step will be to apply their analytical framework to other island-dwelling species, particularly those, such as snails or reptiles, that have less ability to disperse than birds do. These analyses could be further



Figure 1 | Bird biodiversity. The theory of island biogeography^{7,8}, proposed in the 1960s, is a milestone in our understanding of how biodiversity is established and maintained. Valente *et al.*⁶ tested this theory on a global scale using data for island-dwelling birds from 41 archipelagos. Their results confirm key predictions of this theory. **a**, The authors report that two-thirds of the birds native to archipelagos arose from a process of species formation called anagenesis, which typically occurs on isolated islands (those far from the mainland). Anagenesis has given rise to birds such as the Bolle's pigeon (*Columba bollii*) of the Canary Islands. **b**, Another process of species formation, called cladogenesis, is most common on large, isolated islands. The authors report that of the birds they studied, a group called Hawaiian honeycreepers had the greatest number of species (33 in total) that arose by cladogenesis. One example of such species is Hawaii's i'iwi, or scarlet honeycreeper (*Drepanis coccinea*).

informed by incorporating into this approach species' functional traits¹¹, such as body size and diet.

The implications of Valente and colleagues' results extend beyond the field of island biogeography. For example, characterization of the relationship between island area and extinction rate contributes to the discussion in conservation science about how to assess the effects on biodiversity of habitat loss and fragmentation during the Anthropocene (the name proposed for the current phase of planetary history, in which human activity has a dominant influence on the environment). This is relevant to today's world, in which natural habitats are becoming increasingly isolated^{12,13}.

An important aspect of Valente and colleagues' study is their approach of considering an archipelago as a unit, rather than focusing on individual islands. This aligns with the idea¹⁴ that archipelagos might be the most appropriate units in which to frame analyses of biodiversity at large spatial and temporal scales. Analysis of large spatial units in biogeography is not a new approach; however, these units generally take the form of geometric shapes, such as grid squares, that do not directly correspond to ecological boundaries (for example, those defined by vegetation type) and their associated communities. By contrast, archipelagos represent natural units. It is likely that substantial strides will be made in our understanding of island biogeography from further analyses of ecological patterns and processes undertaken at the archipelago scale, especially if geological dynamics are incorporated. To paraphrase E. O. Wilson¹⁵: it is

archipelagos that are “the logical laboratories of biogeography and evolution”.

Kostas A. Triantis is in the Department of Biology, National and Kapodistrian University of Athens, Athens GR-15784, Greece. **Thomas J. Matthews** is at the School of Geography, Earth and Environmental Sciences, and at the Birmingham Institute of Forest Research, University of Birmingham, Birmingham B15 2TT, UK. e-mails: ktriantis@biol.uoa.gr; t.j.matthews@bham.ac.uk

- Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K. & Triantis, K. A. *Science* **357**, eaam8326 (2017).
- Losos, J. B. & Ricklefs, R. E. *Nature* **457**, 830–836 (2009).
- Warren, B. H. *et al.* *Ecol. Lett.* **18**, 200–217 (2015).
- Gillespie, R. G. & Clague, D. A. *Encyclopedia of Islands* (Univ. California Press, 2009).
- Lomolino, M. V. & Brown, J. H. *Q. Rev. Biol.* **84**, 357–390 (2009).
- Valente, L. *et al.* *Nature* **579**, 92–96 (2020).
- MacArthur, R. H. & Wilson, E. O. *Evolution* **17**, 373–387 (1963).
- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
- Valente, L. *et al.* *Curr. Biol.* **27**, 1660–1666 (2017).
- Steadman, D. W. *Extinction and Biogeography of Tropical Pacific Birds* (Univ. Chicago Press, 2006).
- Pigot, A. L. *et al.* *Nature Ecol. Evol.* **4**, 230–239 (2020).
- Haddad, N. M. *et al.* *Sci. Adv.* **1**, e1500052 (2015).
- Russell, J. C. & Kueffer, C. *Annu. Rev. Environ. Resour.* **44**, 31–60 (2019).
- Triantis, K. A., Economo, E. P., Guilhaumon, F. & Ricklefs, R. E. *Glob. Ecol. Biogeogr.* **24**, 594–605 (2015).
- Wilson, E. O. in *The Theory of Island Biogeography Revisited* (eds Losos, J. B. & Ricklefs, R. E.) 1–12 (Princeton Univ. Press, 2010).

This article was published online on 19 February 2020.

Tropical carbon sinks are out of sync

Anja Rammig

A survey of tree establishment, growth and mortality shows that the rate at which Amazonian tropical forests take up carbon dioxide has slowed since the 1990s, whereas signs of a potential slowdown in Africa appeared only in 2010. **See p.80**

The total area of the world that is covered by tropical forest is declining because of deforestation, land degradation and fires – a trend that has increased over the past few years¹. At the same time, human-induced climate change is altering the functioning of tropical forests². During the 1990s and early 2000s, structurally intact tropical forests actively removed carbon from the atmosphere (in the form of carbon dioxide) through photosynthesis, and stored it as biomass. Such forests have been responsible for about 50% of the terrestrial

carbon sink³. Hubau *et al.*⁴ report on page 80 that this globally crucial tropical carbon sink is becoming saturated in both Amazonian and African rainforests, but with different patterns of change.

Forests act as a net carbon sink when the amount of carbon gained through the establishment of new trees and tree growth is larger than the amount lost through tree mortality. In these circumstances, the quantity of carbon stored in the biomass increases over time. The interplay of carbon gains, losses and stocks

determines the period of time for which carbon remains in the forest, which is known as the carbon residence time⁵.

Hubau and colleagues monitored tree establishment, growth and mortality in 244 undisturbed old-growth forest plots in Africa across 11 countries, between 1968 and 2015, and compared their data with similar measurements from 321 plots in Amazonia⁶. Such long-term monitoring is essential for identifying trends and drivers of the carbon sink in forest biomass, but is highly challenging and costly in terms of coordination, labour and funding – particularly in the tropics, where access to field sites is difficult and working conditions are harsh (Fig. 1). The authors find that the carbon sink in African tropical-forest biomass was stable for the 30 years up to 2015, in contrast to the sink in Amazonian tropical forests, for which the annual net amount of accumulated carbon started to decline around 1990 (Fig. 2). So what drives the slowdown of the tropical carbon sink, and why are there differences between Amazonian and African tropical forests?

The authors report a long-term trend of increasing carbon gains in the forests on both continents throughout the period studied, which correlates with the increase in atmospheric CO₂ concentrations. They attribute the



JOÃO MARCOS ROSA

Figure 1 | Taking an inventory of the Amazon rainforest. A researcher takes measurements of a tree trunk at a height of 2 metres above the ground. Long-term monitoring such as this can be used to estimate the amount of carbon stored by tropical forests.

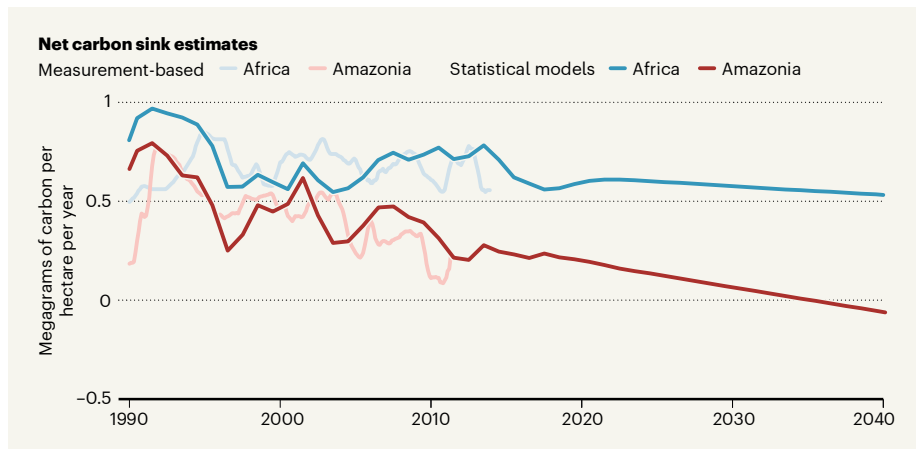


Figure 2 | Estimates and projections of tropical carbon sinks. Hubau *et al.*⁴ have estimated the net amount of carbon that was absorbed from the atmosphere by tropical forests – the tropical carbon sinks – in Africa and Amazonia for the period from 1968 to 2015, using measurements of tree establishment, growth and mortality; only estimates from 1990 onwards are shown. The data show that the sink in Amazonia has declined since the 1990s, whereas the African sink was stable for the 30 years up to 2015. The authors also estimated the carbon sinks using statistical models, which they extrapolated to 2040. The extrapolations suggest that, by 2030, the carbon sink in Africa will be 14% lower than in 2010–15, whereas the Amazonian carbon sink will reach zero by 2035. Data shown are mean values; see Fig. 3 of ref. 4 for confidence intervals.

rising gains to CO₂ fertilization – an increase in carbon uptake by plants that occurs as atmospheric CO₂ levels rise. However, they find that increasing mean annual temperatures and drought since 2000 have reduced tree growth and thus offset the increase in carbon gains, with smaller reductions in Africa than in Amazonia.

Hubau *et al.* go on to show that high carbon gains persisted for longer in Africa than in Amazonia because the warming rate was slower, there were fewer droughts and air temperatures were generally lower (because African forests are located at higher elevations). And, in contrast to an earlier study⁶, the authors were able to clearly attribute the decline of carbon gains in Amazonia to increasing temperatures and repeated extreme drought events, on the basis of a statistical analysis of their data. The researchers find no signs of the CO₂-fertilization effect levelling off on either continent.

Although the authors attribute the decline in carbon gains on both continents to climatic drivers, other limiting factors might be responsible – such as competition between trees for light and nutrients, and the general availability of nutrients on each continent. These factors were not considered in their statistical analysis, but might further constrain tree growth and weaken the sink as atmospheric CO₂ concentrations continue to increase. Such limitations have been hinted at from experiments in which the atmospheric concentration of CO₂ is enriched in a specific area of an ecosystem⁷, but no such experiment has been carried out in highly diverse, old-growth tropical forests such as those in Africa and Amazonia.

In addition to the trends in carbon gains, Hubau *et al.* find that carbon losses in Africa

were stable from the 1990s until a decade ago, and then started to increase. By comparison, carbon losses in Amazonia had already started to increase in the 1990s. This continental difference seems to be because trees in Amazonia grow faster and have shorter carbon residence times than do those in African forests. Carbon dioxide fertilization might increase growth rate and carbon gains, but it also leads to quicker losses – CO₂-fertilized trees grow fast and die young^{5,6}, and therefore might not necessarily contribute to the carbon sink in the long term. The authors find that tree mortality associated with chronic long-term heat

“The authors estimate that the Amazonian carbon sink will reach zero by 2035.”

and drought leads to increased carbon losses, and that this effect is more pronounced in Amazonian than in African tropical forests as a result of accelerated warming rates in Amazonia since 2000. Data from the most intensively monitored African plots indicate that carbon losses in those forests began increasing from about 2010.

The authors extrapolate their statistical models up to the year 2040, and thereby suggest that the carbon sink will decline on both continents. They estimate that, by 2030, the carbon sink in Africa will be 14% lower than in 2010–15, whereas the Amazonian carbon sink will reach zero by 2035 (that is, there will be no net carbon uptake from the atmosphere). These extrapolations need to be interpreted carefully, however, because they are

in striking contrast to projections made by global models – which predict a strong, continuing carbon sink due to CO₂ fertilization in intact tropical forests⁸. Recently reported models⁹ of vegetation growth that consider nutrient cycling show that the Amazonian-forest carbon sink is strongly constrained by the availability of phosphorus in soils. Hubau and colleagues’ findings underline the need to understand other factors that affect tree mortality and forest dynamics, in addition to such nutrient feedbacks, so that these can be integrated into global models².

So, what does a pan-tropical decline of the carbon sink in intact forests imply for the current climate crisis? Calculations of the maximum amount of anthropogenic carbon emissions that can be emitted to limit global warming to well below 2 °C – the goal of the 2015 Paris climate agreement – count on the continuation of a large tropical carbon sink¹⁰. Hubau and co-workers’ finding that tropical sinks are disappearing and could very soon turn into carbon sources suggests that, as well as strong protection of intact tropical forest, even faster reductions of anthropogenic greenhouse-gas emissions than those set out in the agreement will be needed to prevent catastrophic climate changes.

Anja Rammig is at the Technical University of Munich, TUM School of Life Sciences, Weihenstephan, Freising 85354, Germany. e-mail: anja.rammig@tum.de

1. Song, X.-P. *et al.* *Nature* **560**, 639–643 (2018).
2. Trumbore, S., Brando, P. & Hartmann, H. *Science* **349**, 814–818 (2015).
3. Pan, Y. *et al.* *Science* **333**, 988–993 (2011).
4. Hubau, W. *et al.* *Nature* **579**, 80–87 (2020).
5. Körner, C. *Science* **355**, 130–131 (2017).
6. Brienen, R. J. W. *et al.* *Nature* **519**, 344–348 (2015).
7. Norby, R. J. *et al.* *New Phytol.* **209**, 17–28 (2016).
8. Huntingford, C. *et al.* *Nature Geosci.* **6**, 268–273 (2013).
9. Fleischer, K. *et al.* *Nature Geosci.* **12**, 736–741 (2019).
10. Steffen, W. *et al.* *Proc. Natl Acad. Sci. USA* **115**, 8252–8259 (2018).


Single-particle spectroscopy for functional nanomaterials

<https://doi.org/10.1038/s41586-020-2048-8>

Received: 26 February 2019

Accepted: 7 January 2020

Published online: 4 March 2020

 Check for updates

Jiajia Zhou¹✉, Alexey I. Chizhik²✉, Steven Chu^{3,4}✉ & Dayong Jin^{1,5}✉

Tremendous progress in nanotechnology has enabled advances in the use of luminescent nanomaterials in imaging, sensing and photonic devices. This translational process relies on controlling the photophysical properties of the building block, that is, single luminescent nanoparticles. In this Review, we highlight the importance of single-particle spectroscopy in revealing the diverse optical properties and functionalities of nanomaterials, and compare it with ensemble fluorescence spectroscopy. The information provided by this technique has guided materials science in tailoring the synthesis of nanomaterials to achieve optical uniformity and to develop novel applications. We discuss the opportunities and challenges that arise from pushing the resolution limit, integrating measurement and manipulation modalities, and establishing the relationship between the structure and functionality of single nanoparticles.

As was pointed out by Richard Feynman in his classic lecture ‘There’s plenty of room at the bottom’¹, the manipulation of atoms or molecules will allow what we now call nanostructures to be tailored with accuracy and with desirable physical and chemical properties. This requires establishing the relationship between structure and properties for each nanostructure, and observing them with sub-nanometre resolution. Although there is still a long way to go to achieve Feynman’s vision, there has been tremendous progress in nanotechnology and characterization techniques towards realizing the possibilities that he conceived decades ago^{2–7}.

Among the breakthroughs that have stimulated the growth of nanotechnology is the controlled synthesis and characterization of luminescent nanoparticles^{8–15}. The use of single-photon detectors has boosted the development of bright and photostable nanoparticles with tunable emission properties, which have attracted great attention from multiple disciplines^{16–19}. Super-resolution microscopy methods that can resolve single nanoparticles beyond the diffraction limit have become another trigger of progress in nanotechnology²⁰.

Single luminescent nanoparticles, although synthesized from the same batch, can often be heterogeneous in terms of size, shape, defects, surface groups and charges. These are core issues in fundamental research related to material science, crystallography and interfacial chemistry, and are crucial for reproducibility, functionality and applications. Electron microscopy is an essential tool for observing the structure of individual nanoparticles, but it barely accesses their optical functions. Conventional fluorescence spectroscopy can show certain signs of heterogeneity in nanoparticle properties—for instance, ensemble-spectrum broadening or multi-exponentiality of fluorescence decay—but these are derived by ensemble averaging. Single-particle spectroscopy (SPS) is a rapidly developing class of techniques that enable the discernment of individual features of single particles and thus to provide direct information on their heterogeneity. By measuring the heterogeneity

of the optical properties of a single particle, it is possible to analyse the influence of particle size, shape, surface state, composition, geometric orientation and the local environment.

Advances in material synthesis methods and SPS enhance the quality of new types of nanoparticle and the homogeneity of their optical properties. However, the more we pursue perfection in nanoparticle design, the greater the challenges become. With the growth of the complexity of material systems and characterization techniques, future advances in luminescent nanomaterials will be built on interdisciplinary integration of materials sciences, optical imaging and spectroscopy beyond the diffraction limit. This Review focuses on inorganic luminescent nanoparticles, which have been the subject of many studies^{21–23}. We first present new insights that have been gained through SPS techniques. We show the advantages of controlled synthesis towards achieving optical uniformity and field-responsive properties of single nanoparticles, which can be tailored for diverse applications. We discuss the current opportunities and challenges in advancing next-generation SPS.

Luminescent nanoparticles under SPS

Since the first demonstration of single-molecule measurements^{24,25}, rapid progress in optical microscopy has made it possible to routinely see the fluorescence of individual emitters using highly sensitive photodetectors. Advances in single-photon avalanche diodes and electron-multiplied charge-coupled device cameras with high quantum efficiency and low noise have enabled the wide application of single-molecule fluorescence spectroscopy and imaging in commercial and custom-built microscopy systems. For more than a decade, SPS has advanced understanding of the heterogeneity of nanoparticles and facilitated discoveries of the underpinning photophysics. Correlative microscopic SPS methods have been developed in both a passive detection manner, such as in situ atomic-force microscopy (AFM) or

¹Institute for Biomedical Materials and Devices (IBMD), Faculty of Science, University of Technology Sydney, Sydney, New South Wales, Australia. ²Third Institute of Physics, Georg-August University of Göttingen, Göttingen, Germany. ³Department of Physics, Stanford University, Stanford, CA, USA. ⁴Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA. ⁵UTS-SUSTech Joint Research Centre for Biomedical Materials and Devices, Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen, China. ✉e-mail: dayong.jin@uts.edu.au; jiajia.zhou@uts.edu.au; alexey.chizhik@phys.uni-goettingen.de; schu@stanford.edu

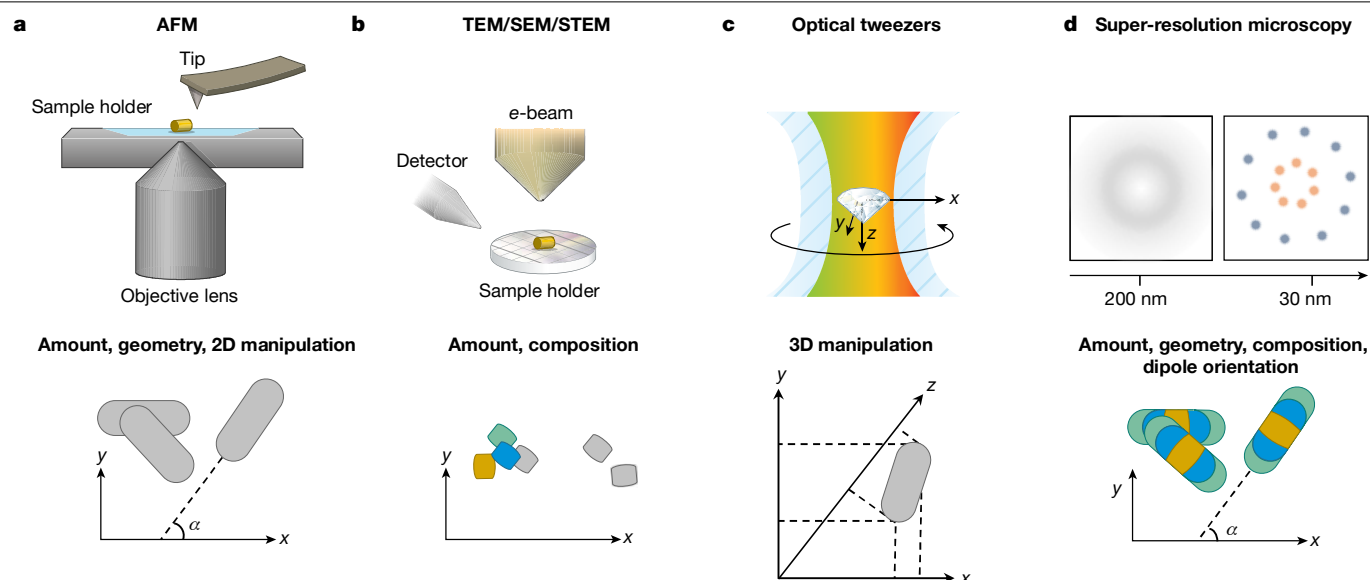


Fig. 1 | Correlative methods providing deterministic information about a single nanoparticle. **a**, In situ AFM integrated with optical microscopy (top) reveals the particle amount, geometry and two-dimensional (2D) orientation (bottom). **b**, TEM/SEM/STEM correlation (top) confirms the particle amount and composition (bottom). **c**, Optical tweezers (top) manipulate the three-

dimensional orientation of a single nanoparticle (bottom) by rotating the polarization of the trapping laser. **d**, Super-resolution microscopy (top) resolves the composition and transition dipole orientation of the nanoparticles, in addition to the particle amount and its two-dimensional orientation (bottom).

ex situ transmission and scanning electron microscopy (TEM and SEM, respectively), and a positive manipulation manner, such as the use of an AFM tip or optical tweezers (Fig. 1). The optical heterogeneity information provided by SPS helps to guide the synthesis of uniform nanoparticles and improve their performance in each application (Fig. 2). In the following, we discuss insights obtained from SPS for different nanoparticles, and exemplify several situations in which the fundamental optical properties of nanoparticles—uniform or heterogeneous—make a difference in applications.

Quantum dots

Quantum dots are artificial atoms with discrete, atomic-like energy levels and a spectrum of narrow transitions. However, populations of different sizes and shapes in an ensemble state induce inhomogeneous spectral shifts and deformation. SPS has been used to decode the discrete nature of excited states of quantum dots since 1996²⁶. In the same year, fluorescence intermittency (blinking) from single CdSe quantum dots was revealed, in which the intensity jumped between on- and off-states under continuous excitation²⁷. Later, it was shown that blinking of quantum dots can have different characteristics¹⁴. The commonly observed (so-called A-type) correlated blinking of fluorescence intensity and modulation of excited-state lifetime are due to charging and discharging of the nanocrystal core. In B-type blinking, changes in emission intensity are not accompanied by modulation in emission dynamics. In thick-shell CdSe/CdS quantum dots, another type of blinking was observed, in which fluctuations of the excited-state lifetime were accompanied by stable emission intensity²⁸. Studies of single-particle fluorescence blinking led to advanced understanding of charge recombination and trapping processes in quantum dots and enabled the design of quantum dots with reduced blinking^{11,29}.

SPS has revealed the relationship between emission polarization of single quantum dots and their aspect ratio³⁰. The dependence of the polarization on the particle shape and orientation of single CdSe/CdS nanorods has been characterized using correlated AFM and optical microscopy (Fig. 1a)³¹. In these experiments, the orientation angles of a single nanorod were identified using AFM as 5°, 125° and 135°, in

good agreement with the emission polarization angles of 7°, 124° and 133°, respectively.

Correlated AFM single-particle fluorescence microscopy confirmed a suboptimal ensemble-fluorescence quantum yield of quantum dots caused by non-emissive particles³². A correlative study of atomic structure, chemical composition and time-resolved single-photon counting of the same single quantum dots revealed inhomogeneity in the quantum yields of single nonblinking giant CdSe/CdS quantum dots³³. In particular, high-angle annular dark-field detection in scanning transmission electron microscopy (STEM) (Fig. 1b), paired with energy-dispersive X-ray spectroscopy, has been used to visualize the distance between the core and the surface of a CdSe/CdS architecture for spectroscopic correlation³³.

The study of heterogeneity has motivated the use of controlled growth to make optically uniform quantum dots, which has a key role in achieving specific performance indicators for applications. For instance, Fan and co-workers³⁴ reported that uniform biaxially strained quantum dots led to a two-fold reduction of the single-particle linewidth compared with hydrostatically strained quantum dots; the former can generate continuous-wave lasing with a low threshold of 8.4 kW cm⁻² (Fig. 2a), whereas the latter can only generate pulsed-laser emissions with a threshold more than two times higher.

Fluorescent nanodiamonds

Fluorescent nanodiamonds that contain colour centres, such as nitrogen-vacancy centres, have been used as single-photon sources since the early 2000s³⁵. SPS is an essential tool in photon antibunching, which is a standard technique for verification of fluorescence from a single quantum emitter. The method exploits the fact that a single emitter emits only one photon upon transition from the excited to the ground state.

SPS has provided many insights into the photostability of nanodiamonds when reducing their size. In 2010, a discrete 5-nm nanodiamond was found to blink and bleach over several hours of illumination¹³. In 2014, the silicon vacancy in a 1.6-nm nanodiamond³⁶ was reported to be chemically stable, but to blink and be optically unstable, with fluorescence lasting only for tens of minutes. Presumably owing to their surface defects, smaller nanodiamonds usually contain a smaller

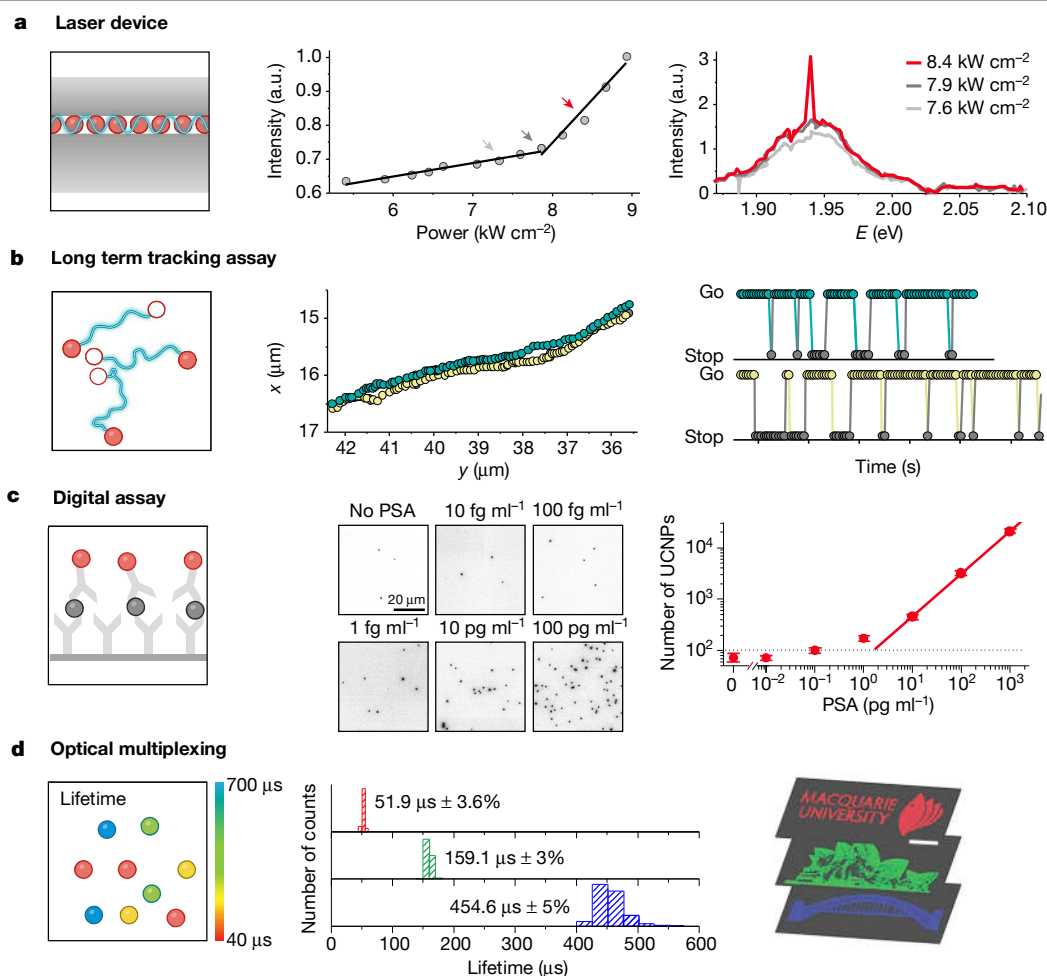


Fig. 2 | Optical uniformity of nanoparticles advances biological and nanophotonics applications. **a**, Left, biaxially strained quantum dots exhibiting narrowed emission linewidth at a single-particle level, facilitating low-threshold laser generation. Middle, normalized integration of the signal emitted from a quantum dot laser as a function of peak power. Right, emission spectra above and below the lasing threshold, corresponding to the points indicated by the arrows in the middle panel. a.u., arbitrary units. Image adapted from ref. ³⁴ (Springer Nature). **b**, Left, use of a single nanodiamond for reliable transport tracking, owing to its superior photostability. Middle, the trajectories of two single fluorescent nanodiamonds. Right, go (movement) phases are shown in green (nanodiamond 1) and yellow

(nanodiamond 2) and stop (static) phases are shown in grey. Image adapted from ref. ³⁸ (Springer Nature). **c**, Left, uniform-brightness single UCNPs enabling the digital assay of biomarkers. Middle, upconversion microscopy images of dilutions of prostate specific antigen (PSA) in 25% serum. Right, detection limit of 1.2 pg ml^{-1} in the digital readout. Error bars indicate the standard deviation from three replicate wells. Image adapted with permission from ref. ⁴⁶ (copyright 2017 American Chemical Society). **d**, Excited-state lifetime measurement of tunable UCNPs enabling time-domain optical multiplexing applications, including data storage and anti-counterfeiting (right). The middle panel shows the lifetime distributions of a series of UCNPs. Image adapted from ref. ⁴⁷ (Springer Nature).

fraction of active emitters and their emission is not uniform in intensity. Current approaches to fabricating nanodiamonds have limited controllability in particle size compared to the wet-chemistry synthesis methods used for other luminescent nanoparticles. There is a need for a synthesis approach that can produce both morphologically and optically uniform nanodiamonds.

Photostable nanoparticles are useful for long-term tracking with high temporal resolution. The single-particle tracking technique can be used to detect minor abnormalities, reveal single-molecule biophysical mechanisms and diagnose diseases at an early stage^{21,37}. Haziza and co-workers³⁸ developed a quantitative assay in mouse hippocampal neurons by tracking single 30-nm nanodiamonds (Fig. 2b). They achieved a temporal resolution of 50 ms and suggested that this assay was sufficiently sensitive to detect a change of protein concentration ($\sim 30\%$). In a control experiment, they demonstrated that random blinking quantum dots could affect the trajectory reconstruction (for example, owing to the existence of dark periods and inferred segments), leading to biased tracking parameters.

Upconversion nanoparticles

Upconversion nanoparticles (UCNPs) are typically (co-)doped by lanthanide ions and represent an emerging type of nonlinear optical material that absorbs low-energy near-infrared photons to produce high-energy emissions in the visible and ultraviolet regions. The single-particle nature of UCNPs is usually confirmed by the correlation between AFM/TEM/SEM (Fig. 1a, b) images and confocal/wide-field fluorescence images^{39–41}. Single UCNPs are photostable for hours^{40,42}. Quantitative measurements of the brightness of individual nanoparticles have revealed that inert-shell passivation could prevent the ‘short-circuit’ effect of the migration of sensitized photon energy to surface quenchers⁴³. Measurements of the fine-splitting emission spectra in individual 8-nm UCNPs have indicated possible heterogeneity of their surface conditions or disorder in the distribution of lanthanides within the nanoparticle⁴⁴.

Because of their nonlinear nature, UCNPs behave very differently in ensemble spectroscopy and in SPS measurements⁴⁵. Ensemble-spectroscopic studies have concluded that optimal doping concentrations

of Tm^{3+} and Er^{3+} should be smaller than 1 mol% and 2 mol% in a $\beta\text{-NaYF}_4$ host, respectively, limited by the concentration quenching effect. However, SPS has revealed that concentration quenching is highly power-dependent^{39,44}. When the excitation power density reaches 10^4 W cm^{-2} or higher, single UCNPs highly doped with 8 mol% Tm^{3+} or 20 mol% Er^{3+} are orders of magnitude brighter than conventional low-doped UCNPs^{39,44}. Recently, a core-shell-shell design of single UCNPs highly doped with 8 mol% Er^{3+} in a NaYbF_4 host has been reported⁴¹ to have high brightness, each emitting about 200 photons per second under a low excitation-power density of 8 W cm^{-2} .

Compared with other fluorescent nanoparticles, UCNPs have better brightness uniformity. Uniform-intensity UCNPs have enabled an ultrasensitive quantification of biomolecules by directly counting the number of single nanoparticles. This type of quantification is known as the single-molecule (digital) upconversion-linked immunosorbent assay (digital ULISA) (Fig. 2c)⁴⁶. The quantification of single molecules is anticipated to be possible once one-to-one binding between the UCNPs and the antibody can be realized.

The excited-state lifetime is a unique quantity with a relatively small deviation compared to other spectroscopic signals, such as emission intensity and peak ratios in UCNPs or other lanthanide-doped nanoparticles. Lifetime uniformity allows one to achieve a single-population distribution of lifetimes at both the single-nanoparticle and ensemble levels^{47,48}, facilitating accurate sensing by using the lifetime modality. On the basis of single-population characteristics, tunable lifetime control is a promising tool for multiplexing applications such as imaging and data storage⁴⁷ (Fig. 2d).

Carbon dots

Carbon dots are a new type of carbon-based luminescent nanoparticle. Carbon dots consist of a carbon crystalline or amorphous core and various luminescent and non-luminescent surface groups, which provide a combination of unique or unexpected optical properties. A growing number of investigations aim to understand the luminescence mechanism of carbon dots. SPS has shown that whereas most of their luminescence properties largely resemble those of organic molecules⁴⁹, the interaction of luminescent surface groups with a carbon core and non-luminescent groups could determine the blinking character of carbon dots^{50,51}. The introduction of specific functional groups to carbon dots was found to not only affect their surface charge to display distinctive colours, but to also direct their selective labelling. Negatively charged green carbon dots were localized within the endosomes/lysosomes while positively charged blue carbon dots were mostly located within the nucleus⁵⁰.

Lead halide perovskite nanoparticles

Lead halide perovskite nanoparticles (LHPs) are colloidal nanocrystals with an APbX_3 -type perovskite lattice, where A is a monovalent cation and X is a halide ion. In recent years, LHPs have attracted an explosive increase in research interest regarding their control synthesis, optical characterization and optoelectronic-device applications²³. SPS has enabled the comprehensive study of the optical performance of LHPs, including their blinking behaviour⁵², exciton dynamics⁵³ and single-photon emission^{54,55}.

Responses to external fields

Apart from intrinsic electronic transitions that enable nanomaterials to display spectacular emissions in different optical dimensions, excitons and electrons can also respond to a stimulus in an external field, such as temperature, magnetic and electrical fields. Their diverse luminescent emissions can 'dance' with the external fields, which not only provides new angles to advance knowledge in their photophysical and material properties, but also enables the development of field-responsive sensors using single nanoparticles^{16,18,56,57}. Statistical measurements of

optical signals from single nanoparticles can identify whether those from the same synthesis batch have synchronous response upon application of the stimulus. Asynchronous responses usually contribute to the study of the heterogeneity and branched properties of nanoparticles. A synchronous response can give universal structure information about nanoparticles, demonstrating their uniformity.

Cryogenic and vacuum conditions

Cryogenic and vacuum conditions can be applied to SPS to reveal exciton behaviours^{58–60}, which provide new insights for decoding emitting states that are complex in atmospheric conditions. By inspecting 120 single CdSe/CdS quantum dots at 300 K in air, at 300 K in vacuum and at 30 K in vacuum, it was found that the nanocrystal was mostly in a charged state under vacuum, and oscillated between a charged state and a neutral state at 300 K in air, where non-radiative Auger recombination was thermally activated (Fig. 3a)⁶¹. The correlated quantum yield and lifetime change suggested that the enhanced Auger recombination was due to electron localization on the nanoparticle surface. Low-temperature SPS has been used to analyse the excited-state lifetime, blinking and polarization of single LHPs to confirm the existence of their bright triplet excitons⁵³.

Magnetic field

Magnetic fields are used in SPS to characterize the Zeeman splitting and Rashba effect of the excited states of single nanoparticles^{62–64}. For example, magneto-SPS has been used to record the evolution of the emission spectra of CsPbBr_3 LHPs as a function of magnetic-field amplitude⁶². By measuring 17 nanoparticles, two subpopulations with different Zeeman splitting evolutions were identified. Among these nanoparticles, five had their z axis nearly parallel to the magnetic field, and the others did not display measurable Zeeman splitting because they were randomly orientated with respect to the magnetic field⁶².

Cryogenic temperature conditions have usually been applied in magnetic-field SPS to minimize non-radiative interference. Very recently, Tamarat and co-workers⁶⁵ observed the direct spectroscopic signature of a dark singlet state from single LHPs in a low-temperature (4 K) magnetic field. By assessing 28 nanoparticles, the singlet state was measured to be several millielectronvolts lower than the bright triplet state. When turning on the magnetic field, they observed an extra redshifted peak and a slow component in the lifetime curve that was associated with the singlet state (Fig. 3b). It was the magnetic field that induced the dark-bright coupling, which enabled relaxation from the triplet to the singlet state, followed by emission from the singlet state. This was further confirmed by measurements of the emission, lifetime and polarization spectra in a variable magnetic field (from 0 T to 7 T).

The responses of single nanoparticles to temperature and magnetic fields can be used to develop nanoscale sensors. The ground-state spin triplet of the nitrogen-vacancy colour centre is sensitive to both a magnetic field and thermally induced lattice strain. These properties underpin nanomagnetometry and temperature sensing using single nitrogen-vacancy-containing nanodiamonds^{18,56,66}. Tools such as the AFM tip (Fig. 1a) and optical tweezers (Fig. 1c) can be used to move single-nanodiamond sensors in two- and three-dimensional spaces^{4,56,67}.

Electric field

The electric field has provided new insights into the excited-state dynamics of single semiconductor nanoparticles. The correlated and uncorrelated fluctuation of fluorescence intensity and lifetime for A- and B-type blinking behaviours in quantum dots have been determined by recording the fluorescence transients from single CdSe/CdS quantum dots with increasing negative potential¹⁴. The quantum-confined Stark effect in single semiconductor quantum dots and nanorods can be used to probe voltage changes at the nanoscale⁶⁸. As shown in Fig. 3c, in contrast to the type-I system (in which conduction- and valence-band minima spatially overlap)⁶⁹, core-shell nanorods with type-II

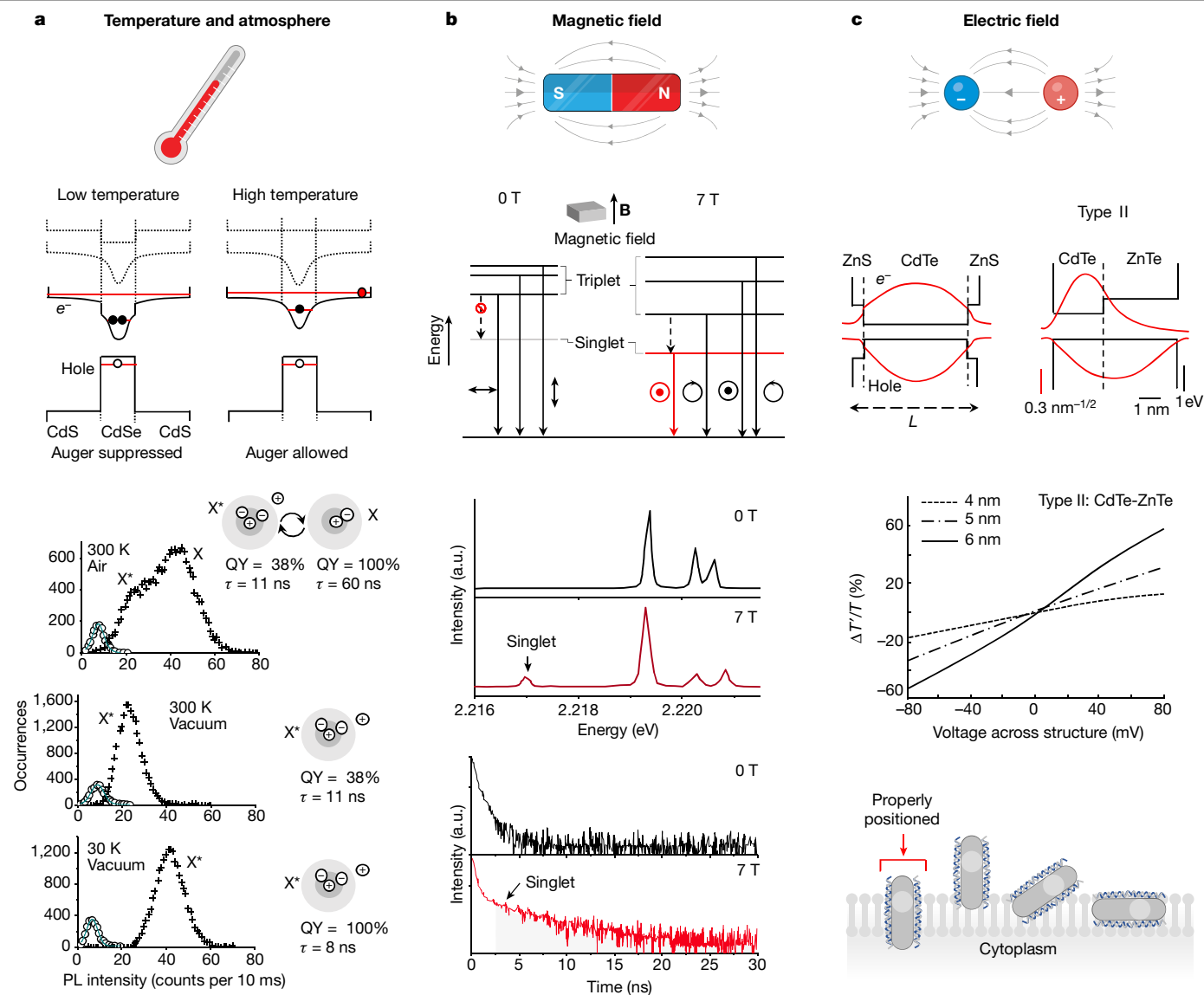


Fig. 3 | Application of external fields to stimulate the response of single nanoparticles dynamically. **a**, Temperature-dependent SPS statistics revealing the thermal activation of non-radiative Auger recombination in charged CdSe/CdS quantum dots, induced by electron delocalization. The schematic shows the band alignment for the negatively charged trion in a CdSe/CdS nanocrystal. Auger processes are suppressed when the trion is in its ground state in a type-I structure (left), but allowed when one of the electrons is delocalized in the shell and interacts with the surface in a quasi-type-II structure (right). Bottom, fluorescence intensity histograms (black crosses) and the corresponding charging states of a quantum dot at 300 K in air (top graph), at 300 K in vacuum (middle graph) and at 30 K in vacuum (bottom graph). The noise intensity is shown by grey open circles and is fitted with a Poisson distribution (blue line). X, neutral state; X^+ , charged state; τ , lifetime; QY, quantum yield. Image adapted from ref. ⁶¹ (Springer Nature). **b**, At 4 K, LHPs

show emission from a dark singlet state with a slow decay rate when the magnetic field strength increases from 0 to 7 T. The schematic shows the Zeeman splitting and polarization of the fine-structure transitions for an LHP with an orthorhombic crystal structure positioned with the z axis parallel to the magnetic field, B , and to the collection axis (Faraday configuration). The three-line spectrum of an LHP nanoparticle at 4 K in zero field changes into a four-line spectrum under a magnetic field (7 T). The fluorescence decay curve under 7 T shows the singlet-state component. Image adapted from ref. ⁶⁵ (Springer Nature). **c**, Type-II semiconductor nanorods exhibit higher voltage-sensing sensitivity than type-I quantum dots, owing to their different band alignment. Quasi-type-II single nanorods have been used for membrane potential sensing. L , length of the nanoparticle. Image adapted with permission from ref. ⁶⁹ (copyright 2013 American Chemical Society) and from ref. ⁷⁰ (American Association for the Advancement of Science).

band alignment (in which conduction- and valence-band minima are spatially separated) showed a stronger field effect because the lower bandgap of the shell allowed the electron to be delocalized in the shell. When the shell could be elongated on one side of the nanoparticle, the field effect became even stronger⁶⁹. Quasi-type-II nanorods have been recently used for membrane potential sensing, and it was found that nanorods need to be aligned perpendicular to the membrane surface for high-sensitivity sensing (Fig. 3c)⁷⁰.

Future opportunities and challenges

Current SPS techniques will continue to advance nanoscale characterizations of the photophysical properties of existing and future materials. Efforts are ongoing to realize new desirable functionalities in smaller and more efficient nanomaterials and to integrate them through both heterogeneous and hybrid designs. This will encourage the development of more powerful SPS-based technologies.

Many biomedical and intracellular applications demand nanoscopic molecular probes and sensors as small as a few nanometres in size. However, for sizes smaller than 10 nm, it is challenging to form uniform nanoparticles and control their morphological and optical properties. The non-uniform size distribution of semiconductor quantum dots will lead to the dispersion of their spectral characteristics. The brightness of emitter-doped nanomaterials, such as nanodiamonds and UCNPs, can be considerably reduced when the volume of each nanoparticle drops. A reduced thickness of passivating shells also degrades the quantum yield of luminescent nanoparticles. These challenges require new methods of controlling the surface chemistry of nanoparticles.

Although highly desirable for many applications, nonlinear upconversion probes do not yet emit luminescence with brightness comparable to that of linear ones. The focus for strong nonlinear upconversion luminescence needs to be on the design of organic–inorganic hybrid materials and proper management of efficient energy transfer, to achieve high efficiencies in photon sensitization, energy transfer and upconversion emission⁷¹. Such hybrid systems demand super-resolution SPS techniques capable of characterizing each step of sophisticated photophysics, particularly at the materials' interface.

The addition of the new functionality of luminescent nanoparticles is anticipated to revolutionize some key scientific and technological areas. For example, the engineering of chirality and its interaction with biomolecules opens up new perspectives in gene analysis⁷². Precisely controlling heterogeneous nanomaterials and/or arbitrarily assembling these building blocks can integrate multiple functionalities into a single nanoparticle. The accuracy of controlling material growth and assembly reaches the nanoscale and even the atomic scale^{73,74}. An interfacial strain-controlled growth method has been reported to form heterogeneous semiconductor nanorods with helical-shell morphology, which induce the development of chirality in quantum dots⁷⁵. The chirality of carbon dots can be engineered by choosing an appropriate chiral surface precursor⁷⁶. Hybrid self-assembly will further integrate diverse functionalities into a nanostructure^{77,78}.

Despite the rapid progress and advances already made in SPS, there are large remaining gaps, from the controlled formation of a diverse range of nanomaterials to the understanding of their fine-tunable photophysical properties. Filling these gaps involves challenges and opportunities for nanoscale and atomic-scale optical characterizations. Here, we discuss eight potential directions in advancing SPS.

Super-resolution SPS

The optical-diffraction limit will continue to constrain resolutions in both the lateral and axial directions in advanced SPS. The solution of super-resolution SPS employs current super-resolution microscopic techniques, which can resolve multiple single nanoparticles and strains that are near each other or localize single emitters within nanomaterials^{5,79}. Hell and co-workers⁸⁰ developed ground-state depletion nanoscopy to resolve 50 ± 8 nm GaInP segments spaced in a GaP nanowire, and found excellent agreement between the ground-state depletion image and the SEM pattern. The recently developed upconversion-stimulated emission depletion microscopy⁷ and its simplified form⁸¹ may be used for decoding the 'barcodes' of heterogeneous UCNPs^{82–85}.

Other super-resolution techniques may be suitable for SPS. Mirror-enhanced axial-narrowing super-resolution can enhance the axial resolution six times by simply replacing the microscope slide with a mirror⁸⁶. A semitransparent metal film on the sample surface can be used to modulate the excited-state lifetime and allows nanometre-axial-resolution measurement⁸⁷. A hybrid method that combines optical and electron microscopy provides a pathway to a resolution down to the single-molecule scale⁸⁸.

The parallel developments of super-resolution techniques and new luminescent nanoparticles mutually enhance each other. SPS advances our understanding and ability to tune the blinking behaviour of single nanoparticles, which has enabled the development of super-resolution optical fluctuation imaging and stochastic optical reconstruction microscopy^{50,89,90}. The interference of scattering patterns of single nanoparticles has been recently employed for high-speed tracking of single molecules in cells⁹¹. Characterizations of the excitation and emission dipoles of single nanoparticles can be used to resolve the orientations of the nanodevice transport process by super-resolution polarization microscopy^{92,93}.

Multi-modal correlative SPS

Correlative microscopy, which integrates SPS techniques, electron microscopy characterization and manipulation methods, combines multi-modal measurements to determine precisely how each of the structural properties of a material determines its overall optical behaviour. Individually, each of these methods provides incomplete, or often even misleading, information.

Recent studies suggest that the luminescence properties associated with carbon dots may also originate from molecular fluorophores or their aggregates that are unavoidably formed during carbon dot synthesis⁹⁴. The co-existence of luminescent molecules in solution is undetectable under TEM. Fluorescence correlation spectroscopy, which provides information on the diffusion of emitters in solution, has allowed researchers to distinguish between nanoparticles and molecular fluorophores⁹⁴.

Care needs to be taken when the measurement substrate also emits luminescence that leads to deceptive results. This is often the case for SPS, because background noise often competes with the signal from individual nanoparticles. Defects in glass, depending on their type and local chemical structure, can emit luminescence covering the whole visible spectrum⁹⁵. What makes SPS measurements more challenging is that optical emissive defects can be created as a result of intense ultraviolet irradiation⁹⁶. The similarity of the optical properties of a single luminescent defect in SiO₂ to those of a dye molecule requires multi-modal correlative SPS to reliably distinguish between them^{95,97}. The combination of measurements that probe the optical and structural properties of emitters can unambiguously attribute fluorescence to a certain origin.

Nanoscale tweezing

Contactless trapping and tweezing of individual nanometre-sized nanoparticles, in combination with SPS, will provide many opportunities for assembling hybrid nanoparticle-based devices and in situ studies of distance- and orientation-dependent phenomena, such as energy transfer and force dynamics between different types of single nanoparticle building blocks. Optical trapping has been applied to confine ultracold atoms⁹⁸ and nanoparticles using a tightly focused laser beam⁹⁹. However, conventional (optical) tweezers face considerable challenges in tweezing nanoparticles⁹⁹, because of the limited trapping force against Brownian motion and thermophoretic forces. New methods involving nanotweezers and nanophotonic devices—for example, microstructure photonic fibre—have been recently developed^{100–102}. The hybrid electrothermoplasmonic nanotweezer enables on-demand, long-range and rapid delivery of single nano-objects to specific plasmonic nanoantennas to achieve two-dimensional assembly¹⁰³. Near-field nanotweezers with specifically designed nanoarchitecture—for example, engineering a bowtie plasmonic aperture at the extremity of a tapered metal-coated optical fibre—could be used to realize three-dimensional optical manipulation of single dielectric objects¹⁰⁴. Other approaches, such as open-access microcavities, may provide both in situ calibration and sensing capabilities during trapping of single nanoparticles¹⁰⁵. Cohen and Moerner^{106,107} developed an

anti-Brownian electrophoretic trap that applies two-dimensional force fields to trap single nanoscale objects in solution.

Surface of single nanoparticles

Non-uniformity of surface species and charge can induce heterogeneity in the optical properties of nanoparticles and affect nearly all their projected applications, from specific molecular targeting to nanodevice self-assembly. SPS has often been insufficient for obtaining chemical information on the surface of a single nanoparticle. Methods such as far-field Raman spectroscopy provide perspective in characterizing the surface species of single nanoparticles. Very recently, Xiong and co-workers¹⁰⁸ developed stimulated Raman excited fluorescence spectroscopy to perform surface molecule analysis with high sensitivity and chemical specificity. They demonstrated all-far-field single-molecule Raman spectroscopy and imaging without the need of plasmonics for near-field enhancement.

The anisotropic surface of nanoparticles determines the heterogeneous binding probabilities of ligands on different crystal facets of a nanoparticle. Surface anisotropy can be resolved by using super-resolution stochastic optical reconstruction microscopy to resolve tip-end-bound dyes on single upconversion nanorods¹⁰⁹.

Light absorption by single nanoparticles

Fluorescence-based characterizations of single nanoparticles with low quantum yield, or quenched nanoparticles that are close to metallic surfaces or in chemical contact with a quencher, are not accessible to single-nanoparticle sensitivity. These 'dark' nanoparticles can be potentially detected by measuring their absorption of light. However, the several-orders-of-magnitude difference between the absorption cross-section of nanometre-sized nanoparticles and the diffraction-limited focal spot makes it nearly impossible to detect the fraction of excitation light that is absorbed by a particle. Sandoghdar and co-workers¹¹⁰ detected the fraction of light that is absorbed by a single semiconductor quantum dot from a noisy background. This was achieved by normalization of the signal, which allows the reduction of laser intensity fluctuations by an order of magnitude. Polarization modulation of excitation light and the use of samples of lower surface roughness have further improved the sensitivity of absorption imaging and led to quantitative values for the absorption cross-section of single molecules under ambient conditions¹¹¹.

Several other advanced methods have been explored for single-particle absorption imaging and even spectroscopy measurements (Fig. 4a). Confining an atom in a radio-frequency Paul trap has achieved absorption imaging of a single laser-cooled atom¹¹². Using a microtoroidal whispering-gallery-mode resonator¹¹³, the resonance shifts of the resonator could be translated into a photothermal sensitivity of tens of picowatts, which is sufficient to measure the absorption spectrum of a single molecule. Photothermal heating of a temperature-sensitive substrate¹¹⁴ has enabled single-particle absorption imaging. By using evanescent waves with a lightpath comparable with the size of nanoparticles in total-internal-reflection microscopy¹¹⁵, the extinction spectroscopy of single particles can be measured. Bawendi and co-workers¹¹⁶ have recently suggested using bow-tie antennas to confine the excitation field within the tiny dimensions of a plasmonic nanostructure. This confinement design can potentially enable further enhancement of the accuracy of single-particle absorption measurements.

Quantum yield of single nanoparticles

The challenge for SPS in measuring absolute quantum yields lies in discerning the number of photons absorbed by a single particle. There are alternative ways to measure quantum yield, which are based on the ratio of the radiative (k_r) to the non-radiative (k_{nr}) transition rate between the excited state(s) to the ground state of a single nanoparticle. Fine-tuning of either the radiative or the non-radiative rate allows one to measure

the total de-excitation rate ($k_r + k_{nr}$), that is, the inverse excited-state lifetime. Whereas the non-radiative rate of a fluorophore is typically determined by its intrinsic properties and local chemical environment, the probability of radiative de-excitation can be tuned by changing the so-called local density of states of the electromagnetic field¹¹⁷.

The radiative-rate change effect has been achieved for fluorophores placed close to a dielectric interface^{118,119}, a sharp tip of a scanning probe microscope¹²⁰, a metallic mirror¹²¹, a metallic nanoparticle¹²², or between two gold nanoparticles¹²³ or silver mirrors of an optical resonator¹²⁴ (Fig. 4b). Although such methods have been applied for organic molecules, the enhanced photostability of nanoparticles enables more precise measurements. A simplified approach has been reported by Brokmann and co-workers¹¹⁸, in which they measured the modulation of the excited-state lifetime of a single quantum dot by placing a droplet of a polymer on top of a substrate. This allowed them to estimate the quantum yield of a single nanocrystal in the bright (on) state to be close to unity.

High-throughput SPS and data analysis

Whereas SPS measurements are usually limited to nanoparticles with sufficient brightness and the SPS method primarily relies on repeated single-particle experiments to achieve statistical results, ensemble methods can be used as a prescreening tool before selecting a smaller number of samples for single-particle analysis. High-throughput SPS and automation of data analysis are desirable for the implementation of single-particle studies into routine sample analysis. A widefield imaging scheme using a commercially available hyperspectral imaging system or a prism to disperse spectral information can drastically enhance the detection throughput and speed^{125,126}. The latter configuration has been reported to capture the emission spectra of about 100 randomly distributed single molecules within snapshots of a few milliseconds.

Machine (deep) learning can go beyond the limits of conventional data analysis^{127–129}. Deep learning has been recently used in analysing single-molecule patterns¹³⁰ and to reconstruct a widefield image into a super-resolution one without substantive implementation once the computational model has been built¹³¹. This indicates that repeated experiments in SPS measurements may become unnecessary by using deep learning to identify and record the optical signatures of single nanoparticles.

SPS standardization

Many ensemble measurements of nanomaterials must be optimized to become quantitative, given that the results obtained by different research groups can be affected by different instrumentation settings and measurement environments. For UCNPs in the ensemble form, the enhancement factors in emission intensity and quantum yield of inert shell-coated UCNP samples can vary from a few to a few orders of magnitude, even when using similar design strategies⁴¹. The quantum yields of UCNPs are strongly dependent on the excitation power and the particle density of UCNPs in powder or suspension.

To make these comparisons realistic, SPS can provide an absolute number of emitted photons at different excitation power densities. Until commercial suppliers can offer a less-complicated optics setup for SPS—for example, a peripheral control package and accessories—a metrology platform should be established to serve the materials science community. Such a standardized platform would be particularly important for nonlinear optical conversion, in which the illumination power density substantially contributes to diverging electronic behaviours^{7,39,44,132}. SPS standardization and accessibility will enable the rapid search for efficient and uniform nanoparticles from various synthesis methods, recipes or experimenters. These nanoparticles can be selected for a variety of potential applications according to their performance in terms of power-dependent intensities, intensity distribution, and saturation in brightness.

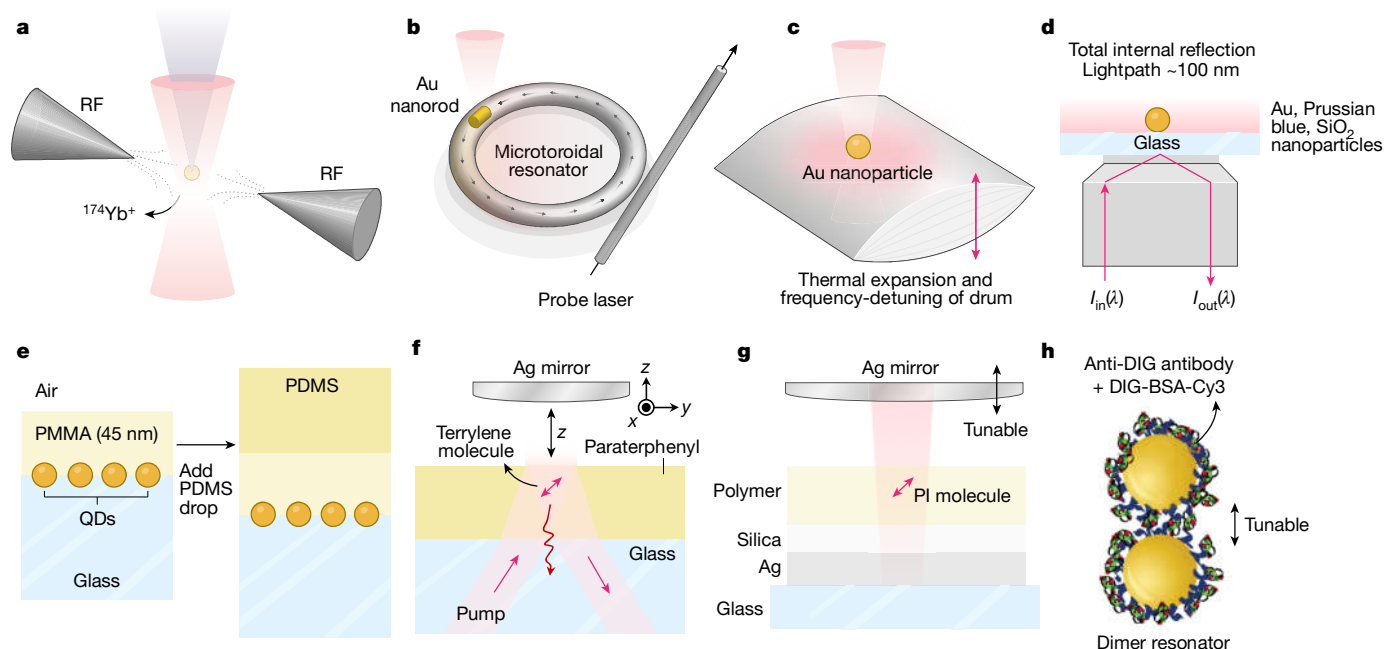


Fig. 4 | Perspective for advanced SPS. **a–h**, Photonics architectures adopted for absorption (**a–d**) and quantum yield (**e–h**) measurements for single objects. **a**, A laser-cooled $^{174}\text{Yb}^+$ ion is confined in a radio frequency (RF) Paul trap formed by the electric quadrupole (dashed lines) between two tungsten needles. Image adapted from ref. ¹¹² (Springer Nature). **b**, A gold nanoparticle on a toroidal microresonator is pumped to generate a photothermal absorption signal, which shifts the toroid resonance frequency. The resonance frequency is probed with a fibre-coupled tunable-frequency laser. Image adapted from ref. ¹¹³ (Springer Nature). **c**, 10-nm gold nanoparticles are measured on a silicon nitride drum under a tensile stress of 20 MPa for absorption imaging. Image adapted from ref. ¹¹⁴ (National Academy of Sciences). **d**, Single-particle extinction spectroscopy by total internal reflection. $I_{\text{in}}(\lambda)$, intensity of input light; $I_{\text{out}}(\lambda)$, intensity of output light; λ , wavelength. Image adapted with permission from ref. ¹¹⁵ (copyright 2019 Wiley-VCH Verlag GmbH & Co. KGaA). **e**, The decay rate of quantum dots is first measured when they are close to a glass–air interface and then measured when a polydimethylsiloxane (PDMS) drop is added to displace the interface far from

the emitter. PMMA, polymethylmethacrylate. Image adapted with permission from ref. ¹¹⁸ (copyright 2004 American Physical Society). **f**, The quantum yield of single terrylene molecules embedded in paraterphenyl is measured by a silver scanning mirror, which is positioned in front of the sample with a three-dimensional piezo stage. Image adapted with permission from ref. ¹²¹ (copyright 2005 American Physical Society). **g**, The quantum yield of single molecules (PI, a perylene derivative) embedded in a polymer layer is measured using a tunable microcavity, which consists of a silver layer, a spacer between the metallic surface and the molecules, and a top adjustable mirror. Image adapted with permission from ref. ¹²⁴ (copyright 2011 American Chemical Society). **h**, A plasmonic nanoresonator composed of two gold nanoparticles can change the spectral shape of the emission of fluorescent molecules. The plasmonic resonance frequency can be tuned by varying the distance between the gold nanoparticles, which allows changing the emission of a molecule. DIG, digoxigenin molecule; BSA, bovine serum albumin; Cy3, cyanine dye molecule. Image adapted with permission from ref. ¹²³ (copyright 2008 American Physical Society).

In summary, developments in materials science and nanophotonics tools now offer the potential to achieve sensitivity and resolution better than the single-nanoparticle level for the characterization of optical nanomaterials. This will enable the controlled use of nanoparticles with optimal properties for a broad range of applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2048-8>.

1. Feynman, R. P. There's plenty of room at the bottom. *Eng. Sci.* **23**, 22–36 (1960).
2. Akimov, A. V. et al. Generation of single optical plasmons in metallic nanowires coupled to quantum dots. *Nature* **450**, 402–406 (2007).
3. Ropp, C. et al. Nanoscale imaging and spontaneous emission control with a single nano-positioned quantum dot. *Nat. Commun.* **4**, 1447 (2013).
4. Geiselmann, M. et al. Three-dimensional optical manipulation of a single electron spin. *Nat. Nanotechnol.* **8**, 175–179 (2013).
5. Rittweger, E., Han, K. Y., Irvine, S. E., Eggeling, C. & Hell, S. W. STED microscopy reveals crystal colour centres with nanometric resolution. *Nat. Photon.* **3**, 144–147 (2009).

This work demonstrated the deterministic trapping and three-dimensional manipulation of single nanodiamonds using optical tweezers.

6. Hanne, J. et al. STED nanoscopy with fluorescent quantum dots. *Nat. Commun.* **6**, 7127 (2015).
7. Liu, Y. et al. Amplified stimulated emission in upconversion nanoparticles for super-resolution nanoscopy. *Nature* **543**, 229–233 (2017).
8. Wang, X., Zhuang, J., Peng, Q. & Li, Y. A general strategy for nanocrystal synthesis. *Nature* **437**, 121–124 (2005).
9. Mahler, B. et al. Towards non-blinking colloidal quantum dots. *Nat. Mater.* **7**, 659–664 (2008).
10. Wang, F. et al. Simultaneous phase and size control of upconversion nanocrystals through lanthanide doping. *Nature* **463**, 1061–1065 (2010).
11. Chen, O. et al. Compact high-quality CdSe–CdS core-shell nanocrystals with narrow emission linewidths and suppressed blinking. *Nat. Mater.* **12**, 445–451 (2013).
12. Chang, Y.-R. et al. Mass production and dynamic imaging of fluorescent nanodiamonds. *Nat. Nanotechnol.* **3**, 284–288 (2008).
13. Bradac, C. et al. Observation and control of blinking nitrogen-vacancy centres in discrete nanodiamonds. *Nat. Nanotechnol.* **5**, 345–349 (2010).
14. Galland, C. et al. Two types of luminescence blinking revealed by spectroelectrochemistry of single quantum dots. *Nature* **479**, 203–207 (2011).
15. McGuinness, L. P. et al. Quantum measurement and orientation tracking of fluorescent nanodiamonds inside living cells. *Nat. Nanotechnol.* **6**, 358–363 (2011).
16. Zhang, Q., Li, Y. L. & Tsien, R. W. The dynamic control of kiss-and-run and vesicular reuse probed with single nanoparticles. *Science* **323**, 1448–1453 (2009).
17. Maletinsky, P. et al. A robust scanning diamond sensor for nanoscale imaging with single nitrogen-vacancy centres. *Nat. Nanotechnol.* **7**, 320–324 (2012).
18. Kucsko, G. et al. Nanometre-scale thermometry in a living cell. *Nature* **500**, 54–58 (2013).

19. Tatebayashi, J. et al. Room-temperature lasing in a single nanowire with quantum dots. *Nat. Photon.* **9**, 501–505 (2015).
20. Wöll, D. & Flors, C. Super-resolution fluorescence imaging for materials science. *Small Methods* **1**, 1700191 (2017).
21. Jin, D. et al. Nanoparticles for super-resolution microscopy and single-molecule tracking. *Nat. Methods* **15**, 415–423 (2018).
22. Himmelstoß, S. F. & Hirsch, T. A critical comparison of lanthanide based upconversion nanoparticles to fluorescent proteins, semiconductor quantum dots, and carbon dots for use in optical sensing and imaging. *Methods Appl. Fluoresc.* **7**, 022002 (2019).
23. Akkerman, Q. A., Raino, G., Kovalenko, M. V. & Manna, L. Genesis, challenges and opportunities for colloidal lead halide perovskite nanocrystals. *Nat. Mater.* **17**, 394–405 (2018).
24. Moerner, W. E. & Kador, L. Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.* **62**, 2535–2538 (1989).
25. Orrit, M. & Bernard, J. Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal. *Phys. Rev. Lett.* **65**, 2716–2719 (1990).
26. Empedocles, S. A., Norris, D. J. & Bawendi, M. G. Photoluminescence spectroscopy of single CdSe nanocrystallite quantum dots. *Phys. Rev. Lett.* **77**, 3873–3876 (1996).
27. Nirmal, M., Dabbousi, B. O., Bawendi, M. G. & Macklin, J. Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature* **383**, 802 (1996).
28. Galland, C. et al. Lifetime blinking in nonblinking nanocrystal quantum dots. *Nat. Commun.* **3**, 908 (2012).
29. Efros, A. L. & Nesbitt, D. J. Origin and control of blinking in quantum dots. *Nat. Nanotechnol.* **11**, 661–671 (2016).
30. Hu, J. et al. Linearly polarized emission from colloidal semiconductor quantum rods. *Science* **292**, 2060–2063 (2001).
31. Hadar, I., Hitin, G. B., Sitt, A., Faust, A. & Banin, U. Polarization properties of semiconductor nanorod heterostructures: from single particles to the ensemble. *J. Phys. Chem. Lett.* **4**, 502–507 (2013).
32. Ebenstein, Y., Mokari, T. & Banin, U. Fluorescence quantum yield of CdSe/ZnS nanocrystals investigated by correlated atomic-force and single-particle fluorescence microscopy. *Appl. Phys. Lett.* **80**, 4033–4035 (2002).
33. Orfield, N. J. et al. Quantum yield heterogeneity among single nonblinking quantum dots revealed by atomic structure-quantum optics correlation. *ACS Nano* **10**, 1960–1968 (2016).
34. Fan, F. et al. Continuous-wave lasing in colloidal quantum dot solids enabled by facet-selective epitaxy. *Nature* **544**, 75–79 (2017).
- This study demonstrated that strain-optimized quantum dots with narrow emission linewidth can achieve continuous-wave lasing under a low pumping threshold.**
35. Beveratos, A., Brouri, R., Gacoin, T., Poizat, J.-P. & Grangier, P. Nonclassical radiation from diamond nanocrystals. *Phys. Rev. A* **64**, 061802 (2001).
36. Vlasov, I. I. et al. Molecular-sized fluorescent nanodiamonds. *Nat. Nanotechnol.* **9**, 54–58 (2014).
37. Zeng, X. et al. Visualization of intra-neuronal motor protein transport through upconversion microscopy. *Angew. Chem. Int. Ed.* **58**, 9262–9268 (2019).
38. Haziza, S. et al. Fluorescent nanodiamond tracking reveals intraneuronal transport abnormalities induced by brain-disease-related genetic risk factors. *Nat. Nanotechnol.* **12**, 322–328 (2017).
- This work experimentally demonstrated the advantage of using photostable nanodiamond to perform long-term single-particle tracking.**
39. Zhao, J. et al. Single-nanocrystal sensitivity achieved by enhanced upconversion luminescence. *Nat. Nanotechnol.* **8**, 729–734 (2013).
40. Wu, S. et al. Non-blinking and photostable upconverted luminescence from single lanthanide-doped nanocrystals. *Proc. Natl Acad. Sci. USA* **106**, 10917–10921 (2009).
- This paper reported non-blinking and non-bleaching fluorescence from single lanthanide-doped nanocrystals.**
41. Liu, Q. et al. Single upconversion nanoparticle imaging at sub-10 W cm⁻² irradiance. *Nat. Photon.* **12**, 548–553 (2018).
- This work developed a type of UCNP with uniform and bright emissions at low-power irradiance.**
42. Park, Y. I. et al. Nonblinking and nonbleaching upconverting nanoparticles as an optical imaging nanoprobe and T1 magnetic resonance imaging contrast agent. *Adv. Mater.* **21**, 4467–4471 (2009).
43. Ma, C. et al. Optimal sensitizer concentration in single upconversion nanocrystals. *Nano Lett.* **17**, 2858–2864 (2017).
44. Gargas, D. J. et al. Engineering bright sub-10-nm upconverting nanocrystals for single-molecule imaging. *Nat. Nanotechnol.* **9**, 300–305 (2014).
45. Zhou, J., Xu, S., Zhang, J. & Qiu, J. Upconversion luminescence behavior of single nanoparticles. *Nanoscale* **7**, 15026–15036 (2015).
46. Farka, Z., Mickert, M. J., Hlavacek, A., Skladal, P. & Gorris, H. H. Single molecule upconversion-linked immunosorbent assay with extended dynamic range for the sensitive detection of diagnostic biomarkers. *Anal. Chem.* **89**, 11825–11830 (2017).
47. Lu, Y. et al. Tunable lifetime multiplexing using luminescent nanocrystals. *Nat. Photon.* **8**, 32–36 (2014).
- This paper first reported the controllable growth of a library of lifetime poly-dispersed UCNP for optical multiplexing.**
48. Fan, Y. et al. Lifetime-engineered NIR-II nanoparticles unlock multiplexed in vivo imaging. *Nat. Nanotechnol.* **13**, 941–946 (2018).
49. Ghosh, S. et al. Photoluminescence of carbon nanodots: dipole emission centers and electron-phonon coupling. *Nano Lett.* **14**, 5656–5661 (2014).
50. Chizhik, A. M. et al. Super-resolution optical fluctuation bio-imaging with dual-color carbon nanodots. *Nano Lett.* **16**, 237–242 (2016).
51. Khan, S. et al. Charge-driven fluorescence blinking in carbon nanodots. *J. Phys. Chem. Lett.* **8**, 5751–5757 (2017).
52. Tian, Y. et al. Giant photoluminescence blinking of perovskite nanocrystals reveals single-trap control of luminescence. *Nano Lett.* **15**, 1603–1608 (2015).
53. Becker, M. A. et al. Bright triplet excitons in caesium lead halide perovskites. *Nature* **553**, 189–193 (2018).
54. Hu, F. et al. Superior optical properties of perovskite nanocrystals as single photon emitters. *ACS Nano* **9**, 12410–12416 (2015).
55. Park, Y. S., Guo, S., Makarov, N. S. & Klimov, V. I. Room temperature single-photon emission from individual perovskite quantum dots. *ACS Nano* **9**, 10386–10393 (2015).
56. Balasubramanian, G. et al. Nanoscale imaging magnetometry with diamond spins under ambient conditions. *Nature* **455**, 648–651 (2008).
57. Rendler, T. et al. Optical imaging of localized chemical events using programmable diamond quantum nanosensors. *Nat. Commun.* **8**, 14701 (2017).
58. Tessier, M. D., Javaux, C., Maksimovic, I., Lorient, V. & Dubertret, B. Spectroscopy of single CdSe nanoplatelets. *ACS Nano* **6**, 6751–6758 (2012).
59. Labeau, O., Tamarat, P. & Lounis, B. Temperature dependence of the luminescence lifetime of single CdSe/ZnS quantum dots. *Phys. Rev. Lett.* **90**, 257404 (2003).
60. Rainò, G. et al. Single cesium lead halide perovskite nanocrystals at low temperature: fast single-photon emission, reduced blinking, and exciton fine structure. *ACS Nano* **10**, 2485–2490 (2016).
61. Javaux, C. et al. Thermal activation of non-radiative Auger recombination in charged colloidal nanocrystals. *Nat. Nanotechnol.* **8**, 206–212 (2013).
62. Fu, M. et al. Neutral and charged exciton fine structure in single lead halide perovskite nanocrystals revealed by magneto-optical spectroscopy. *Nano Lett.* **17**, 2895–2901 (2017).
63. Isarov, M. et al. Rashba effect in a single colloidal CsPbBr₃ perovskite nanocrystal detected by magneto-optical measurements. *Nano Lett.* **17**, 5020–5026 (2017).
64. Cannesson, D. et al. Negatively charged and dark excitons in CsPbBr₃ perovskite nanocrystals revealed by high magnetic fields. *Nano Lett.* **17**, 6177–6183 (2017).
65. Tamarat, P. et al. The ground exciton state of formamidinium lead bromide perovskite nanocrystals is a singlet dark state. *Nat. Mater.* **18**, 717–724 (2019).
- This paper reported the direct spectroscopic signature of dark exciton emission from single lead bromide perovskite nanocrystals at cryogenic temperatures and under magnetic fields.**
66. Maze, J. R. et al. Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* **455**, 644–647 (2008).
67. Neukirch, L. P., von Haartman, E., Rosenholm, J. M. & Nick Vamvakas, A. Multi-dimensional single-spin nano-optomechanics with a levitated nanodiamond. *Nat. Photon.* **9**, 653–657 (2015).
68. Park, K., Deutsch, Z., Li, J. J., Oron, D. & Weiss, S. Single molecule quantum-confined Stark effect measurements of semiconductor nanoparticles at room temperature. *ACS Nano* **6**, 10013–10023 (2012).
69. Marshall, J. D. & Schnitzer, M. J. Optical strategies for sensing neuronal voltage using quantum dots and other semiconductor nanocrystals. *ACS Nano* **7**, 4601–4609 (2013).
70. Park, K. et al. Membrane insertion of—and membrane potential sensing by—semiconductor voltage nanosensors: feasibility demonstration. *Sci. Adv.* **4**, e1601453 (2018).
71. Wen, S. et al. Future and challenges for hybrid upconversion nanosystems. *Nat. Photon.* **13**, 828–838 (2019).
72. Sun, M. et al. Site-selective photoinduced cleavage and profiling of DNA by chiral semiconductor nanoparticles. *Nat. Chem.* **10**, 821–830 (2018).
73. Tan, C., Chen, J., Wu, X.-J. & Zhang, H. Epitaxial growth of hybrid nanostructures. *Nat. Rev. Mater.* **3**, 17089 (2018).
74. Laramy, C. R., O'Brien, M. N. & Mirkin, C. A. Crystal engineering with DNA. *Nat. Rev. Mater.* **4**, 201–224 (2019).
75. Ji, B. et al. Strain-controlled shell morphology on quantum rods. *Nat. Commun.* **10**, 2 (2019).
76. Đorđević, L. et al. Design principles of chiral carbon nanodots help convey chirality from molecular to nanoscale level. *Nat. Commun.* **9**, 3442 (2018).
77. Liu, W. et al. Fluorescent nanodiamond–gold hybrid particles for multimodal optical and electron microscopy cellular imaging. *Nano Lett.* **16**, 6236–6244 (2016).
78. Li, X., Zhao, D. & Zhang, F. Multifunctional upconversion-magnetic hybrid nanostructured materials: synthesis and bioapplications. *Theranostics* **3**, 292–305 (2013).
79. Kianinia, M. et al. All-optical control and super-resolution imaging of quantum emitters in layered materials. *Nat. Commun.* **9**, 874 (2018).
80. Oracz, J. et al. Ground state depletion nanoscopy resolves semiconductor nanowire barcode segments at room temperature. *Nano Lett.* **17**, 2652–2659 (2017).
81. Chen, C. et al. Multi-photon near-infrared emission saturation nanoscopy using upconversion nanoparticles. *Nat. Commun.* **9**, 3290 (2018).
82. Fischer, S., Swaback, J. K. & Alivisatos, A. P. Controlled isotropic and anisotropic shell growth in β-NaLnF₄ nanocrystals induced by precursor injection rate. *J. Am. Chem. Soc.* **139**, 12325–12332 (2017).
83. Zhuo, Z. et al. Manipulating energy transfer in lanthanide-doped single nanoparticles for highly enhanced upconverting luminescence. *Chem. Sci.* **8**, 5050–5056 (2017).
84. Liu, D. et al. Three-dimensional controlled growth of monodisperse sub-50 nm heterogeneous nanocrystals. *Nat. Commun.* **7**, 10254 (2016).
85. Zhang, Y. et al. Multicolor barcoding in a single upconversion crystal. *J. Am. Chem. Soc.* **136**, 4893–4896 (2014).
86. Yang, X. et al. Mirror-enhanced super-resolution microscopy. *Light Sci. Appl.* **5**, e16134 (2016).
87. Chizhik, A. I., Rother, J., Gregor, I., Janshoff, A. & Enderlein, J. Metal-induced energy transfer for live cell nanoscopy. *Nat. Photon.* **8**, 124 (2014).
88. Prigozhin, M. B. et al. Bright sub-20-nm cathodoluminescent nanoprobe for electron microscopy. *Nat. Nanotechnol.* **14**, 420–425 (2019).
89. Watanabe, T. M., Fukui, S., Jin, T., Fujii, F. & Yanagida, T. Real-time nanoscopy by using blinking enhanced quantum dots. *Biophys. J.* **99**, L50–L52 (2010).
90. Yang, X. et al. Versatile application of fluorescent quantum dot labels in super-resolution fluorescence microscopy. *ACS Photonics* **3**, 1611–1618 (2016).
91. Taylor, R. W. et al. Interferometric scattering microscopy reveals microsecond nanoscopic protein motion on a live cell membrane. *Nat. Photon.* **13**, 480–487 (2019).

92. Zhanghao, K. et al. Super-resolution dipole orientation mapping via polarization demodulation. *Light Sci. Appl.* **5**, e16166 (2016).
93. Wang, M. et al. Polarization-based super-resolution imaging of surface-enhanced Raman scattering nanoparticles with orientational information. *Nanoscale* **10**, 19757–19765 (2018).
94. Khan, S. et al. Small molecular organic nanocrystals resemble carbon nanodots in terms of their properties. *Chem. Sci.* **9**, 175–180 (2018).
95. Chizhik, A. M. et al. Imaging and spectroscopy of defect luminescence and electron–phonon coupling in single SiO₂ nanoparticles. *Nano Lett.* **9**, 3239–3244 (2009).
96. Tarpani, L. et al. Photoactivation of luminescent centers in single SiO₂ nanoparticles. *Nano Lett.* **16**, 4312–4316 (2016).
97. Chizhik, A. I. et al. Measurement of vibrational modes in single SiO₂ nanoparticles using a tunable metal resonator with optical subwavelength dimensions. *Phys. Rev. Lett.* **109**, 223902 (2012).
98. Chu, S. Nobel lecture: the manipulation of neutral particles. *Rev. Mod. Phys.* **70**, 685–706 (1998).
99. Maragò, O. M., Jones, P. H., Gucciardi, P. G., Volpe, G. & Ferrari, A. C. Optical trapping and manipulation of nanostructures. *Nat. Nanotechnol.* **8**, 807–819 (2013).
100. Lin, L. et al. Opto-thermoelectric nanotweezers. *Nat. Photon.* **12**, 195–201 (2018).
101. Crozier, K. B. Quo vadis, plasmonic optical tweezers? *Light Sci. Appl.* **8**, 35 (2019).
102. Xin, H. et al. Single upconversion nanoparticle–bacterium cotrapping for single-bacterium labeling and analysis. *Small* **13**, 1603418 (2017).
103. Ndukaife, J. C. et al. Long-range and rapid transport of individual nano-objects by a hybrid electrothermoplasmonic nanotweezer. *Nat. Nanotechnol.* **11**, 53–59 (2016).
104. Berthelot, J. et al. Three-dimensional manipulation with scanning near-field optical nanotweezers. *Nat. Nanotechnol.* **9**, 295–299 (2014).
105. Trichet, A. et al. Nanoparticle trapping and characterization using open microcavities. *Nano Lett.* **16**, 6172–6177 (2016).
106. Cohen, A. E. Control of nanoparticles with arbitrary two-dimensional force fields. *Phys. Rev. Lett.* **94**, 118102 (2005).
107. Cohen, A. E. & Moerner, W. E. Method for trapping and manipulating nanoscale objects in solution. *Appl. Phys. Lett.* **86**, 093109 (2005).
108. Xiong, H. et al. Stimulated Raman excited fluorescence spectroscopy and imaging. *Nat. Photon.* **13**, 412–417 (2019).
- This work developed an all-far-field single-molecule Raman spectroscopy and imaging technique.**
109. Ren, W. et al. Anisotropic functionalization of upconversion nanoparticles. *Chem. Sci.* **9**, 4352–4358 (2018).
110. Kukura, P., Celebrano, M., Renn, A. & Sandoghdar, V. Imaging a single quantum dot when it is dark. *Nano Lett.* **9**, 926–929 (2009).
111. Celebrano, M., Kukura, P., Renn, A. & Sandoghdar, V. Single-molecule imaging by optical absorption. *Nat. Photon.* **5**, 95 (2011); correction **12**, 309 (2018).
112. Streed, E. W., Jechow, A., Norton, B. G. & Kielpinski, D. Absorption imaging of a single atom. *Nat. Commun.* **3**, 933 (2012).
113. Heylman, K. D. et al. Optical microresonators as single-particle absorption spectrometers. *Nat. Photon.* **10**, 788–795 (2016).
114. Chien, M. H., Brameshuber, M., Rossboth, B. K., Schutz, G. J. & Schmid, S. Single-molecule optical absorption imaging by nanomechanical photothermal sensing. *Proc. Natl Acad. Sci. USA* **115**, 11150–11155 (2018).
115. Li, M. et al. Total internal reflection-based extinction spectroscopy of single nanoparticles. *Angew. Chem. Int. Ed.* **58**, 572–576 (2019).
116. Jensen, R. A. et al. Optical trapping and two-photon excitation of colloidal quantum dots using bowtie apertures. *ACS Photonics* **3**, 423–427 (2016).
117. Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.* **69**, 37–38 (1946).
118. Brokmann, X., Coolen, L., Dahan, M. & Hermier, J. P. Measurement of the radiative and nonradiative decay rates of single CdSe nanocrystals through a controlled modification of their spontaneous emission. *Phys. Rev. Lett.* **93**, 107403 (2004).
119. Macklin, J. J., Trautman, J. K., Harris, T. D. & Brus, L. E. Imaging and time-resolved spectroscopy of single molecules at an interface. *Science* **272**, 255–258 (1996).
120. Ambrose, W. P., Goodwin, P. M., Keller, R. A. & Martin, J. C. Alterations of single molecule fluorescence lifetimes in near-field optical microscopy. *Science* **265**, 364–367 (1994).
121. Buchler, B., Kalkbrenner, T., Hettich, C. & Sandoghdar, V. Measuring the quantum efficiency of the optical emission of single radiating dipoles using a scanning mirror. *Phys. Rev. Lett.* **95**, 063003 (2005).
122. Holzmeister, P. et al. Quantum yield and excitation rate of single molecules close to metallic nanostructures. *Nat. Commun.* **5**, 5356 (2014).
123. Ringler, M. et al. Shaping emission spectra of fluorescent molecules with single plasmonic nanoresonators. *Phys. Rev. Lett.* **100**, 203002 (2008).
124. Chizhik, A. I. et al. Probing the radiative transition of single molecules with a tunable microresonator. *Nano Lett.* **11**, 1700–1703 (2011).
125. Gonell, F. et al. Aggregation-induced heterogeneities in the emission of upconverting nanoparticles at the submicron scale unfolded by hyperspectral microscopy. *Nanoscale Adv.* **1**, 2537–2545 (2019).
126. Zhang, Z., Kenny, S. J., Hauser, M., Li, W. & Xu, K. Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved super-resolution microscopy. *Nat. Methods* **12**, 935–938 (2015).
127. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
128. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
129. Zhou, J., Huang, B., Yan, Z. & Bünzli, J.-C. G. Emerging role of machine learning in light-matter interaction. *Light Sci. Appl.* **8**, 84 (2019).
130. Zhang, P. et al. Analyzing complex single-molecule emission patterns with deep learning. *Nat. Methods* **15**, 913–916 (2018).
131. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36**, 460–468 (2018).
132. Tian, B. et al. Low irradiance multiphoton imaging with alloyed lanthanide nanocrystals. *Nat. Commun.* **9**, 3082 (2018).

Acknowledgements We acknowledge support from the Australian Research Council (ARC) Discovery Early Career Researcher Award Scheme (DE180100669), Shenzhen Science and Technology Program (KQTD20170810110913065) and Australia China Science and Research Fund Joint Research Centre for POCT (ACSRF65827).

Author contributions All authors developed the scope and focus of the Review and contributed to the writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.J., J.Z., A.I.C. or S.C.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2020

Strange-metal behaviour in a pure ferromagnetic Kondo lattice

<https://doi.org/10.1038/s41586-020-2052-z>

Received: 30 July 2019

Accepted: 11 December 2019

Published online: 4 March 2020

 Check for updates

Bin Shen^{1,7}, Yongjun Zhang^{1,7}, Yashar Komijani², Michael Nicklas³, Robert Borth³, An Wang¹, Ye Chen¹, Zhiyong Nie¹, Rui Li¹, Xin Lu¹, Hanoh Lee^{1,6}, Michael Smidman^{1,✉}, Frank Steglich^{1,3}, Piers Coleman^{2,4,✉} & Huiqiu Yuan^{1,5,✉}

A wide range of metals exhibit anomalous electrical and thermodynamic properties when tuned to a quantum critical point (QCP), although the origins of such strange metals have posed a long-standing mystery. The frequent association of strange metals with unconventional superconductivity and antiferromagnetic QCPs^{1–4} has led to the belief that they are highly entangled quantum states⁵. By contrast, ferromagnets are regarded as an unlikely setting for strange metals, because they are weakly entangled and their QCPs are often interrupted by competing phases or first-order phase transitions^{6–8}. Here we provide evidence that the pure ferromagnetic Kondo lattice^{9,10} CeRh₆Ge₄ becomes a strange metal at a pressure-induced QCP. Measurements of the specific heat and resistivity under pressure demonstrate that the ferromagnetic transition is continuously suppressed to zero temperature, revealing a strange-metal behaviour around the QCP. We argue that strong magnetic anisotropy has a key role in this process, injecting entanglement in the form of triplet resonating valence bonds into the ordered ferromagnet. We show that a singular transformation in the patterns of the entanglement between local moments and conduction electrons, from triplet resonating valence bonds to Kondo-entangled singlet pairs at the QCP, causes a jump in the Fermi surface volume—a key driver of strange-metallic behaviour. Our results open up a direction for research into ferromagnetic quantum criticality and establish an alternative setting for the strange-metal phenomenon. Most importantly, strange-metal behaviour at a ferromagnetic QCP suggests that quantum entanglement—not the destruction of antiferromagnetism—is the common driver of the varied behaviours of strange metals.

Quantum materials that are augmented by strong electronic correlations are promising for various applications, but the electronic interactions that empower these materials challenge our understanding. One of the most pressing questions in strongly correlated electronic systems is the origin of the strange-metal behaviour that develops at a quantum critical phase transition between a delocalized Fermi liquid, and a localized or partially localized electronic phase. A prime example is the strange-metal behaviour that develops in the normal state of copper oxide superconductors at optimal doping, characterized by a robust linear resistivity and a logarithmic temperature dependence of the specific-heat coefficient^{2,3}; similar behaviour is also observed in various quantum critical heavy electron materials. The underlying universality of strange-metal behaviour that develops in the vicinity of QCPs is currently a subject of intense theoretical interest. One of the valuable ways of identifying the key ingredients of strange-metal behaviour is through experiments that explore new classes of quantum materials.

Kondo lattice systems, which have periodically arranged atoms hosting localized *f* electrons show a rich variety of properties, owing to competition between magnetic interactions among local moments and their magnetic screening by conduction electrons, the so-called Kondo effect¹. The small energy scales of these interactions leads to highly tunable ground states, which is ideal for studying strange-metal behaviour. In a variety of systems, tuning this competition leads to a continuous suppression of antiferromagnetic order at a QCP⁴. However, the outcome when a ferromagnetic (FM) transition is suppressed by a non-thermal tuning parameter is generally different⁷. FM QCPs are usually avoided, owing to the occurrence of a first-order transition¹¹, the intersection of antiferromagnetic phases^{12,13}, or a Kondo cluster glass phase¹⁴. This raises the question of whether antiferromagnetic correlations are crucial for realizing strange-metal behaviour.

Early theoretical studies of itinerant ferromagnets^{6,8} in the framework of Hertz–Millis–Moriya theory¹⁵ predicted that quantum phase transitions in these materials inevitably become first-order as a consequence

¹Center for Correlated Matter and Department of Physics, Zhejiang University, Hangzhou, China. ²Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, USA. ³Max Planck Institute for Chemical Physics of Solids, Dresden, Germany. ⁴Hubbard Theory Consortium, Department of Physics, Royal Holloway, University of London, Egham, UK. ⁵Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing, China. ⁶Present address: Center for Quantum Materials and Superconductivity, Department of Physics, Sungkyunkwan University, Suwon, South Korea. ⁷These authors contributed equally: Bin Shen, Yongjun Zhang. ✉e-mail: msmidman@zju.edu.cn; coleman@physics.rutgers.edu; hqyuan@zju.edu.cn

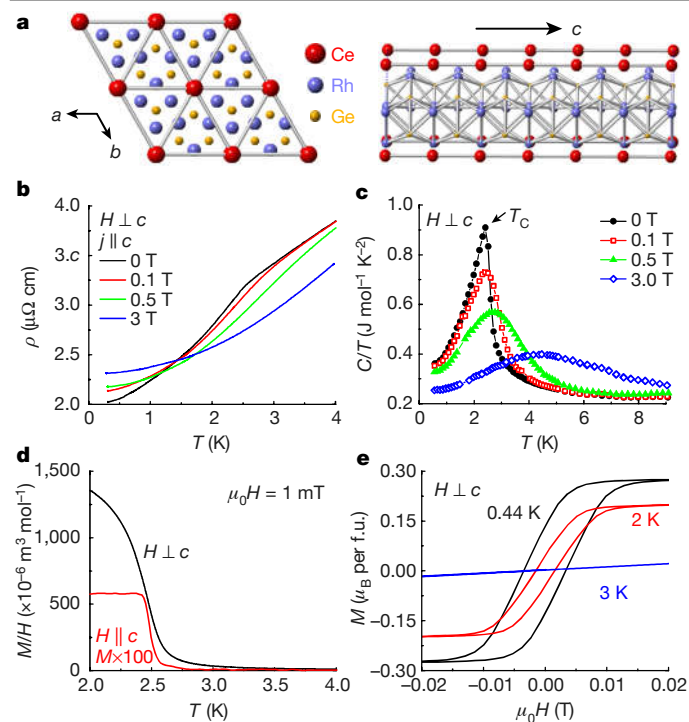


Fig. 1 | Crystal structure and physical properties of CeRh₆Ge₄ at ambient pressure. **a**, Crystal structure of CeRh₆Ge₄. The red, blue and yellow atoms denote Ce, Rh, and Ge, respectively. Left, the structure perpendicular to the *a*–*b* plane, where the Ce atoms have a hexagonal arrangement. Right, the structure perpendicular to the chain direction (*c* axis). **b**, **c**, The resistivity $\rho(T)$ (with the current *j* parallel to the *c* axis; **b**) and specific heat as C/T (**c**) versus *T* for CeRh₆Ge₄, in zero field and various fields applied in the *a*–*b* plane. **d**, Temperature dependence of the magnetization of CeRh₆Ge₄, as M/H , in a field of 1 mT applied both along the *c* axis and in the *a*–*b* plane, where the data for the *c* axis field are scaled by a factor of 100. **e**, Low-field magnetization loops for fields within the *a*–*b* plane at three temperatures. Below T_C , these exhibit hysteresis loops typical of FM order, whereas at 3 K no hysteresis is observed. f.u., formula unit.

of interactions between the critically scattered electron fields, thereby interrupting the development of quantum criticality. However, the recent discovery of an FM QCP in the heavy-fermion system YbNi₄P₂ when tuned by chemical pressure¹⁶ raised the possibility that the FM QCP in these systems is governed by a different universality class involving a breakdown of Kondo screening^{17–19}. The negative pressure required to reach the FM QCP of YbNi₄P₂ necessarily involves chemical doping of the stoichiometric compound, which introduces disorder, complicating the theoretical interpretation. Disorder suppresses first-order transitions⁶, as in the case of ZrZn₂, in which early experiments suggested the presence of an FM QCP²⁰, but improved sample quality led to a first-order transition²¹. Therefore, although the experimental data on YbNi₄P₂ suggest the existence of FM QCPs, definitive proof of such behaviour in a quantum ferromagnet requires using hydrostatic instead of chemical pressure. Cerium-based heavy-fermion ferromagnets, in which pressure can cleanly tune the system to a QCP, are ideally suited for such studies.

CeRh₆Ge₄ is a heavy-fermion ferromagnet with a Curie temperature¹⁰ $T_C = 2.5$ K. The crystal structure (Fig. 1a) consists of triangular lattices of cerium stacked along the *c* axis⁹. The Ce–Ce separation is much smaller along the *c* axis (3.86 Å) than in the triangular planes (7.15 Å), suggesting a quasi-one-dimensional nature to the magnetism. Under hydrostatic pressure, we find that the FM transition of CeRh₆Ge₄ is smoothly suppressed to zero temperature, reaching a QCP at $p_c = 0.8$ GPa.

The temperature dependence of the resistivity $\rho(T)$ and the specific heat (as $C(T)/T$) of single-crystalline CeRh₆Ge₄ both show transition

anomalies at around $T_C \approx 2.5$ K (Fig. 1b, c). When magnetic fields are applied within the *a*–*b* plane, the transition becomes a broadened crossover, consistent with FM ordering. The low-temperature magnetization divided by the applied field, M/H , is shown in Fig. 1d. Measurements up to 300 K demonstrate that the magnetic easy direction lies within the *a*–*b* plane (Extended Data Fig. 1). On cooling, just above T_C the in-plane M/H undergoes a marked enhancement, typical of FM order. For fields along the *c* axis, M/H abruptly increases at the transition. Magnetization loops below T_C for in-plane fields show hysteresis that is characteristic of FM materials (Fig. 1e). $M(H)$ increases rapidly at low fields, reaching $0.28\mu_B$ per Ce atom for $\mu_0 H = 0.017$ T at 0.44 K (μ_B , Bohr magneton; μ_0 , magnetic constant). Upon further increasing the field, there is no hysteresis between up and down field sweeps, and $M(H)$ changes slowly, indicating that $0.28\mu_B$ per Ce atom corresponds to the ordered moment (Extended Data Fig. 1).

The zero-field resistivity and specific-heat coefficient at various pressures are displayed in Fig. 2a, b (see also Extended Data Figs. 3, 4). The evolution of the properties with pressure (*p*) and the resulting *T*–*p* phase diagram are presented in Fig. 3a, b. At T_C the resistivity changes from a linear *T*-dependent behaviour at high temperature to a T^2 -dependent behaviour at low temperatures (Extended Data Fig. 3), where $C(T)/T$ becomes temperature independent. The FM transition, which is suppressed almost linearly by pressure, cannot be detected beyond $p_c = 0.8$ GPa. In the paramagnetic phase above p_c , the aforementioned low-*T* properties of a Fermi liquid are again observed (Extended Data Figs. 3, 4). The temperature at which this Fermi-liquid behaviour onsets (T_{FL}) increases almost linearly with pressure (Fig. 3b). Both the value of the low-temperature $C(T)/T$ and the *A* coefficient of the resistivity in $\rho(T) = \rho_0 + AT^2$ (ρ_0 , residual resistivity) show an incipient divergence when approaching p_c from the FM or paramagnetic side (Fig. 3a). On both Fermi-liquid sides of the phase diagram, the Kadowaki–Woods ratio A/γ^2 (γ , Sommerfeld coefficient) is 1.49×10^{-6} at ambient pressure and $1.33 \times 10^{-6} \mu\Omega \text{ cm mol}^2 \text{ K}^2 \text{ mJ}^{-2}$ at 1.12 GPa, which are close to the value for a 4*f*-electron ground-state degeneracy $N = 4$.

At $p_c = 0.8$ GPa, the resistivity is strictly linear in temperature over two orders of magnitude down to at least 40 mK, whereas $C(T)/T \propto \log(T^*/T)$ over nearly an order of magnitude with $T^* = 2.3$ K (T^* is a characteristic temperature of the spin fluctuation energies)⁴; see Fig. 2c. At 60 mK, $C(T)/T$ reaches a very large value of $1.1 \text{ J mol}^{-1} \text{ K}^{-2}$. Between the FM and paramagnetic phases, there is a fan-shaped strange-metal region with properties similar to canonical antiferromagnetic quantum critical systems such as CeCu_{6-x}Au_x²² and YbRh₂Si₂²³. The pressure dependencies of *A* and γ (Fig. 3a) follow the residual resistivity ρ_0 , which also develops a maximum at p_c , reflecting the presence of quantum critical fluctuations (Extended Data Fig. 3).

At first glance, the strange-metal properties of CeRh₆Ge₄ might be attributed to itinerant quantum criticality, because, aside from the absence of a first-order phase transition, Hertz–Millis–Moriya theory predicts a logarithmic Sommerfeld coefficient and a *T*-linear electron scattering rate, naively equivalent to a *T*-linear resistivity⁴. However, the scattering off long-wavelength FM fluctuations does not relax electron currents, and once this effect is included, $\rho(T)$ is expected to follow a $T^{5/3}$ dependence at low temperature^{4,11}. A *T*-linear resistivity suggests large-angle scattering, a feature typical of local fluctuations involving a wide range of momenta. Moreover, the strength of the logarithmic divergence in the specific-heat coefficient, from fitting $C(T)$ with $(S_0/T^*) \log(T^*/T)$, shows that a large fraction of the local moment entropy, $S_0 \approx (1/10)R \log 2$ (where *R* is the gas constant), is released over a temperature scale T^* (ref. 4). By contrast, the itinerant Hertz–Millis–Moriya theory predicts $S_0 \propto (q_0/q_F)^3$ where q_0 is the momentum cutoff of the itinerant magnetic fluctuations and q_F is the Fermi momentum (Supplementary Information). Applying this theory to the data then requires $q_0 \approx q_F$, which, by Fourier's theorem, implies that the critical spin fluctuations are local. Together with the absence of a first-order phase transition, these features provide strong evidence in favour of a local QCP.

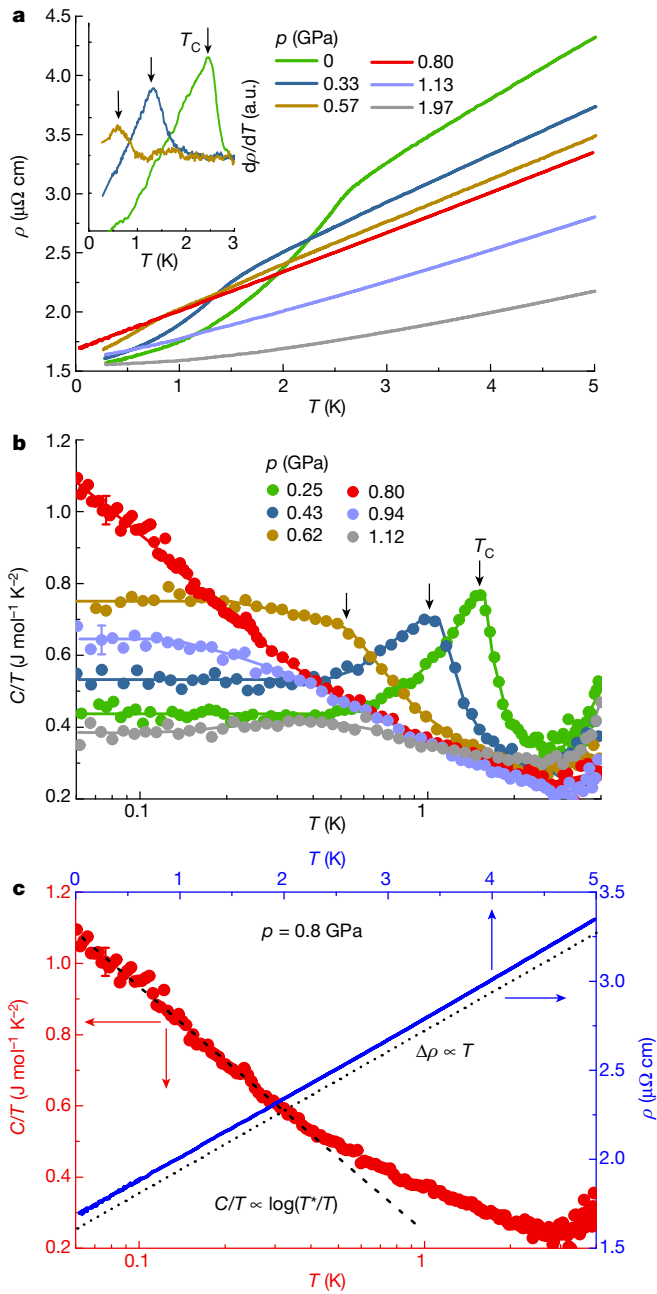


Fig. 2 | Pressure evolution of ferromagnetism in CeRh₆Ge₄ and strange-metal behaviour at the QCP. a, Resistivity of CeRh₆Ge₄ under various hydrostatic pressures. The FM transition is suppressed by pressure, and is no longer observed at $p_c = 0.8$ GPa (red line). Inset, derivative of $\rho(T)$ at lower pressures; the peak position corresponds to T_c . **b**, Specific heat of CeRh₆Ge₄ under hydrostatic pressures. The bulk FM transition is suppressed with pressure, as indicated by the vertical arrows showing the position of T_c . For clarity, not all the data points are displayed. The error bars shown are representative of the scattering of the data at low temperature. A transition to Fermi-liquid behaviour at low temperatures can be observed on either side of p_c , where $C(T)/T$ flattens. **c**, $\rho(T)$ and $C(T)/T$ at $p_c = 0.8$ GPa. $\rho(T)$ exhibits linear behaviour extending from 5 K down to at least 40 mK (dotted line), whereas $C(T)/T$ continues to increase with decreasing temperature, exhibiting a dependence proportional to $\log(T^*/T)$.

In antiferromagnetic heavy-electron metals, the development of T -linear resistivity at the lowest temperatures coincides with an abrupt jump in the Fermi surface volume, accompanied by singular charge fluctuations^{24–26}. It has been argued that such a jump in the Fermi surface

is caused by an abrupt transformation in the pattern of spin entanglement⁵, as the Kondo singlets transform into resonating valence bonds (RVBs) in the spin fluid. This poses a problem, because the spins in a simple ferromagnet are not entangled, which would imply a continuous evolution of the Fermi surface²⁷. As shown below, a clue to unravelling this puzzle comes from the unusual aspect that CeRh₆Ge₄ develops a strange-metal phase at an FM QCP, similar to that observed for the non-stoichiometric material YbNi₄P_{2–x}As_x¹⁶.

Apart from their quasi-one-dimensional nature, a common feature of these two materials is an easy-plane anisotropy. In such systems, the magnetic-order parameter is no longer conserved and will develop marked zero-point fluctuations, which are probably responsible for the severely reduced magnetic moment. This can be seen clearly in a two-site example where the magnetization is perpendicular to the quantization (z) axis of the spins. The ordered phase is a product state that can be expanded in terms of triplets,

$$\frac{|\uparrow_i\rangle + |\downarrow_i\rangle}{\sqrt{2}} \frac{|\uparrow_j\rangle + |\downarrow_j\rangle}{\sqrt{2}} = \frac{|\uparrow_i\uparrow_j\rangle + |\downarrow_i\downarrow_j\rangle}{2} + \frac{1}{\sqrt{2}} \frac{|\uparrow_i\downarrow_j\rangle + |\downarrow_i\uparrow_j\rangle}{\sqrt{2}} \quad (1)$$

where i and j are site indices. An easy-plane anisotropy projects out the equal-spin pairs on the right-hand side, creating a triplet valence bond. In a lattice, the same effect creates a quantum superposition of triplet pairs, forming a triplet RVB state, $|\text{tRVB}\rangle$. Hence, easy-plane anisotropy in FM systems has the same role as magnetic frustration in antiferromagnetic systems, injecting a macroscopic entanglement into the ground state. This leads us to hypothesize that the strange-metal behaviour at the FM QCP has its origins in the magnetic anisotropy.

To test these ideas, we have studied a simplified Kondo lattice model with nearest-neighbour FM couplings with easy-plane anisotropy of the form $-J_{xy}^{ij}(S_i^x S_j^x + S_i^y S_j^y) - J_z^{ij} S_i^z S_j^z$ on a tetragonal lattice, consisting of spin chains along the c direction with weak inter-chain couplings (see Supplementary Information). Here, $S_j^{x,y,z}$ are the x , y and z components of the spin at site j , and J_{xy}^{ij} and J_z^{ij} are the magnetic couplings between the spins at sites i and j . When the chains are weakly coupled, our simulations indicate the development of a second-order phase transition, whereas at higher couplings a first-order phase transition develops. This feature is in agreement with the current observations of FM QCPs developing in quasi-one-dimensional systems. We assume $J_{xy} > J_z$, which has a dual effect: it converts the model into an easy-plane x - y ferromagnet, and generates triplet RVBs. Also, the anisotropy changes the magnetic dispersion at low momenta from quadratic to linear (see Supplementary Information). By switching on the Kondo screening^{26–28} we can then tune the model to the QCP.

Our calculations take advantage of a Schwinger-boson representation of the magnetic moments, which enables us to examine both the magnetic and Kondo-screened parts of the phase diagram, and the QCP that links them together (Fig. 3c). The key feature of this approach is a representation of the spins as bosonic spinons, enabling a dynamical description of the Kondo effect in which neutral local moments fractionalize into negatively charged electrons, leaving behind positively charged Kondo singlets. In the ordered phase, a majority of the moments are aligned, although some form triplet RVB pairs with their neighbours. In an isotropic ferromagnet, the continuous growth of magnetization away from the QCP indicates a continuous change in the fraction of Kondo-screened moments, or a continuous evolution of the Fermi surface. However, when the moments entangled within tRVB states are abruptly released into the Fermi sea, we find (see Supplementary Information) that there is a jump in the Fermi surface volume. The resulting QCP is a plasma in which the Kondo singlets, the electrons and the RVB bonds are in a state of critical dynamical equilibrium, giving rise to singular spin and charge fluctuations as well as a specific-heat coefficient that is logarithmic in temperature (Supplementary Information), in agreement with our experimental results.

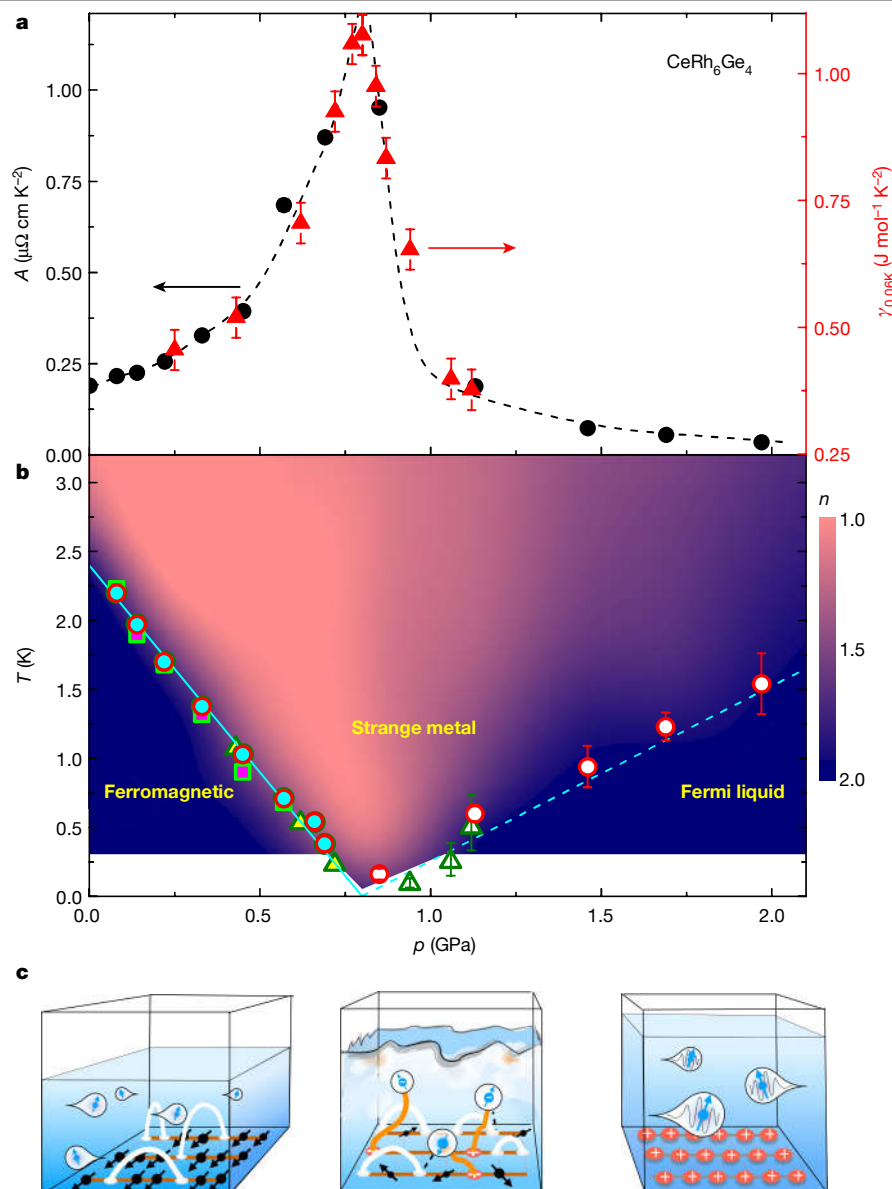


Fig. 3 | Phase diagram of $CeRh_6Ge_4$ under pressure. a, Pressure dependence of the A coefficient of the T^2 term from the resistivity and Sommerfeld coefficient γ (as C/T at 60 mK), which shows a pronounced maximum near the QCP. The error bars for the A coefficient are smaller than the symbols. For γ , the errors correspond to the scattering of the low- T data. The dashed line is a guide to the eye. **b**, T - p phase diagram of $CeRh_6Ge_4$. The circles, triangles and squares for pressures below p_c denote T_c derived from the resistivity, specific heat (d.c. method), and a.c. heat capacity (Extended Data Fig. 5), respectively. The corresponding symbols above p_c mark T_{FL} , below which Fermi-liquid behaviour occurs. The FM transition is suppressed by pressure until the system reaches a QCP at $p_c \approx 0.8$ GPa. Below T_c , and at higher pressures below T_{FL} , Fermi-liquid ground states develop. The colours denote the exponent of $\rho(T)$ calculated as

$n = d(\log(\rho - \rho_0))/d(\log T)$, where the Fermi-liquid states with $n=2$ are dark blue, and the strange-metal phase near the QCP with $n=1$ is shown in pink. **c**, Schematic representation of different phases. In the ordered phase (left), most of the spins are ordered in the plane, although some have RVB bonds. The Fermi surface is small, as represented by the volume of the conduction sea. In the paramagnetic Fermi-liquid phase (right), all the spins are 'ionized' to form heavy electrons that expand the Fermi sea. A background of positively charged singlets are left behind. At the QCP (centre), the system is in a dynamical critical equilibrium in which the moments are fluctuating and the Kondo screening by the conduction electron competes with RVBs for the entanglement. In this region, critical fluctuations strongly scatter the conduction electrons.

Our findings of a pressure-induced QCP in $CeRh_6Ge_4$ demonstrate that an FM system can develop a continuous quantum phase transition in the absence of disorder, a result that at present can only be understood in the framework of local quantum criticality, where Kondo screening is suppressed to zero at the QCP. The observation of strange-metal behaviour at finite temperatures above the QCP—that is, T -linear resistivity and a specific-heat coefficient that is logarithmically divergent in T —expands the scope of this phenomenon to encompass ferromagnets. Central to the strange-metal behaviour in a ferromagnet is a small abrupt jump in the Fermi surface volume. An

experimental observation of such a jump would be an unambiguous test of Kondo breakdown, because there is no unit-cell doubling at an FM phase transition.

Finally, spin-triplet superconducting pairing states have been proposed in FM heavy-fermion systems such as UGe_2 ²⁹ and $URhGe$ ³⁰. Although there is no sign of superconductivity in $CeRh_6Ge_4$ down to 40 mK, it is probable that at sufficiently low temperatures the triplet RVB states that are already present in the critical regime will migrate into the conduction band as a triplet superconducting condensate.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2052-z>.

1. Gegenwart, P., Si, Q. & Steglich, F. Quantum criticality in heavy-fermion metals. *Nat. Phys.* **4**, 186–197 (2008).
2. Daou, R. et al. Linear temperature dependence of resistivity and change in the Fermi surface at the pseudogap critical point of a high- T_c superconductor. *Nat. Phys.* **5**, 31–34 (2008).
3. Legros, A. et al. Universal T -linear resistivity and Planckian dissipation in overdoped cuprates. *Nat. Phys.* **15**, 142–147 (2019).
4. Stewart, G. R. Non-Fermi-liquid behavior in d - and f -electron metals. *Rev. Mod. Phys.* **73**, 797–855 (2001).
5. Senthil, T., Vojta, M. & Sachdev, S. Weak magnetism and non-Fermi liquids near heavy-fermion critical points. *Phys. Rev. B* **69**, 035111 (2004).
6. Belitz, D., Kirkpatrick, T. R. & Vojta, T. First-order transitions and multicritical points in weak itinerant ferromagnets. *Phys. Rev. Lett.* **82**, 4707–4710 (1999).
7. Brando, M., Belitz, D., Grosche, F. M. & Kirkpatrick, T. R. Metallic quantum ferromagnets. *Rev. Mod. Phys.* **88**, 025006 (2016).
8. Chubukov, A. V., Pépin, C. & Rech, J. Instability of the quantum critical point of itinerant ferromagnets. *Phys. Rev. Lett.* **92**, 147003 (2004).
9. Vosswinkel, D., Niehaus, O., Rodewald, U. C. & Pöttgen, R. Bismuth flux growth of CeRh_2Ge_4 and CeRh_2Ge_2 single crystals. *Z. Naturforsch. B* **67**, 1241–1247 (2012).
10. Matsuoka, E. et al. Ferromagnetic transition at 2.5 K in the hexagonal Kondo-lattice compound CeRh_2Ge_4 . *J. Phys. Soc. Jpn* **84**, 073704 (2015).
11. Pfleiderer, C., McMullan, G. J., Julian, S. R. & Lonzarich, G. G. Magnetic quantum phase transition in MnSi under hydrostatic pressure. *Phys. Rev. B* **55**, 8330–8338 (1997).
12. Süllow, S., Aronson, M. C., Rainford, B. D. & Haen, P. Doniach phase diagram, revisited: from ferromagnet to Fermi liquid in pressurized CeRu_2Ge_2 . *Phys. Rev. Lett.* **82**, 2963–2966 (1999).
13. Brando, M. et al. Logarithmic Fermi-liquid breakdown in NbFe_2 . *Phys. Rev. Lett.* **101**, 026401 (2008).
14. Westerkamp, T. et al. Kondo-cluster-glass state near a ferromagnetic quantum phase transition. *Phys. Rev. Lett.* **102**, 206404 (2009).
15. Sachdev, S. *Quantum Phase Transitions* 2nd edn (Cambridge Univ. Press, 2011).
16. Steppke, A. et al. Ferromagnetic quantum critical point in the heavy-fermion metal $\text{YbNi}_4(\text{P}_{1-x}\text{As}_x)_2$. *Science* **339**, 933–936 (2013).
17. Custers, J. et al. The break-up of heavy electrons at a quantum critical point. *Nature* **424**, 524–527 (2003).
18. Schröder, A. et al. Onset of antiferromagnetism in heavy-fermion metals. *Nature* **407**, 351–355 (2000).
19. Yamamoto, S. J. & Si, Q. Metallic ferromagnetism in the Kondo lattice. *Proc. Natl Acad. Sci. USA* **107**, 15704–15707 (2010).
20. Smith, T. F., Mydosh, J. A. & Wohlfarth, E. P. Destruction of ferromagnetism in ZrZn_2 at high pressure. *Phys. Rev. Lett.* **27**, 1732–1735 (1971).
21. Uhlarz, M., Pfleiderer, C. & Hayden, S. M. Quantum phase transitions in the itinerant ferromagnet ZrZn_2 . *Phys. Rev. Lett.* **93**, 256404 (2004).
22. Löhneysen, H. v. et al. Non-Fermi-liquid behavior in a heavy-fermion alloy at a magnetic instability. *Phys. Rev. Lett.* **72**, 3262–3265 (1994).
23. Trovarelli, O. et al. YbRh_2Si_2 : pronounced non-Fermi-liquid effects above a low-lying magnetic phase transition. *Phys. Rev. Lett.* **85**, 626–629 (2000).
24. Paschen, S. et al. Hall-effect evolution across a heavy-fermion quantum critical point. *Nature* **432**, 881–885 (2004).
25. Shishido, H., Settai, R., Harima, H. & Ōnuki, Y. A drastic change of the Fermi surface at a critical pressure in CeRhIn_5 : dHvA study under pressure. *J. Phys. Soc. Jpn* **74**, 1103–1106 (2005).
26. Komijani, Y. & Coleman, P. Emergent critical charge fluctuations at the Kondo breakdown of heavy fermions. *Phys. Rev. Lett.* **122**, 217001 (2019).
27. Komijani, Y. & Coleman, P. Model for a ferromagnetic quantum critical point in a 1D Kondo lattice. *Phys. Rev. Lett.* **120**, 157206 (2018).
28. Wang, J., Chang, Y.-Y., Mou, C.-Y., Kirchner, S. & Chung, C.-H. Quantum phase transition in a two-dimensional Kondo–Heisenberg model: a Schwinger-boson large- N approach. Preprint at <http://arxiv.org/abs/1901.10411> (2019).
29. Saxena, S. S. et al. Superconductivity on the border of itinerant-electron ferromagnetism in UGe_2 . *Nature* **406**, 587–592 (2000).
30. Lévy, F., Sheikin, I., Grenier, B. & Huxley, A. D. Magnetic field-induced superconductivity in the ferromagnet URhGe . *Science* **309**, 1343–1346 (2005).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Crystal growth and characterization

Needle-shaped single crystals of CeRh_6Ge_4 were grown using a bismuth flux⁹. The elements were combined in a molar ratio of Ce:Rh:Ge:Bi of 1:6:4:150, and sealed in an evacuated quartz tube. The tube was heated and held at 1,100 °C for 10 h, then cooled at 3 °C per hour to 500 °C. The tube was then removed, and centrifuged to remove the excess bismuth. The orientation of the crystals was determined using single-crystal X-ray diffraction, and the chemical composition was confirmed using energy-dispersive X-ray spectroscopy. The samples measured under pressure had typical values of $\rho_0 \approx 1.6 \mu\Omega \text{ cm}$ and a residual-resistance ratio of $\text{RRR} = \rho(300 \text{ K})/\rho(0.3 \text{ K}) \approx 45$ (Extended Data Fig. 2).

Physical property measurements

Magnetization measurements were performed using a Magnetic Property Measurement System (Quantum Design). The heat capacity at ambient pressure was measured down to 0.4 K in applied magnetic fields up to 14 T, using a Quantum Design Physical Property Measurement System (PPMS) with a ^3He insert, using the standard relaxation method. Specific-heat experiments under pressure were carried out using a copper beryllium piston-cylinder-type pressure cell³¹. The sample and a piece of lead as pressure gauge were put in a teflon capsule together with Fluorinert serving as a liquid pressure-transmitting medium. The capsule was then mounted inside the pressure cell. The heat capacity of the whole assembly was determined by a compensated heat-pulse method in a dilution refrigerator (Oxford Instruments) down to temperatures of 60 mK. To obtain the heat capacity of the sample the addenda has been recorded in a separate measurement run and subtracted for each pressure from the data obtained of the whole setup including the sample. The pressure inside the cell was determined by the pressure-induced shift of the superconducting transition temperature of the piece of lead, which was measured in a Magnetic Property Measurement System (Quantum Design). The magnetic field was removed in an oscillating fashion to reduce the remanent field (<3 Oe) of the superconducting magnet. The remaining effect on the superconducting transition temperature was compensated for by determining the

shift of the superconducting transition of the lead inside the pressure cell with respect to a reference piece fixed to the outside. Electrical transport and a.c. calorimetry measurements under pressure were carried out in a piston-cylinder clamp-type cell with Daphne oil 7373 as a pressure-transmitting medium. The pressure was also determined from the superconducting transition of lead. The resistivity was measured using the four-contact configuration between 0.04 K and 300 K. The measurements were performed down to 1.9 K, 0.4 K and 0.04 K in a PPMS, ^3He refrigerator and dilution refrigerator, respectively. Data obtained from these measurements are all consistent.

Data availability

All the data supporting the findings are available from the corresponding authors upon reasonable request.

31. Nicklas, M. in *Strongly Correlated Systems: Experimental Techniques* (eds Mancini, F. & Avella, A.) 180, 173–204 (Springer, 2015).

Acknowledgements We thank C. Krellner and M. Brando for discussions, G. Cao and Z. Wang for assisting with ^3He -SQUID measurements and X. Xiao for assistance with single-crystal X-ray diffraction. This work was supported by the National Key R&D Program of China (grants 2017YFA0303100, 2016YFA0300202), the National Natural Science Foundation of China (grants U1632275, 11974306), the Science Challenge Project of China (grant number TZ2016004) and the National Science Foundation of the United States of America, grant number DMR-1830707.

Author contributions H.Y. conceived the study and led the project. The crystals were grown by Y.Z. and H.L. Measurements of the properties at ambient pressure, as well as measurements of the electrical resistivity and a.c. specific heat under pressure, were performed by B.S., Y.Z., A.W., Y.C., Z.N., R.L., X.L. and H.Y. The quasi-adiabatic specific-heat measurements under hydrostatic pressure were measured by R.B. and M.N. The experimental data were analysed by B.S., Y.Z., M.N., H.L., M.S., F.S. and H.Y. Theoretical calculations were performed by Y.K. and P.C. The manuscript was written by Y.K., M.S., F.S., P.C. and H.Y. All authors participated in discussions.

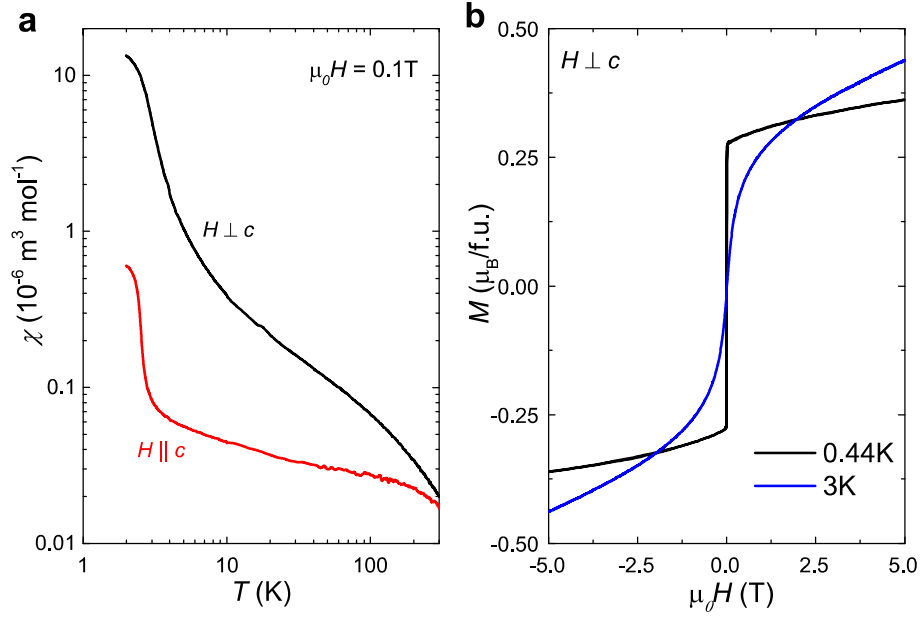
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2052-z>.

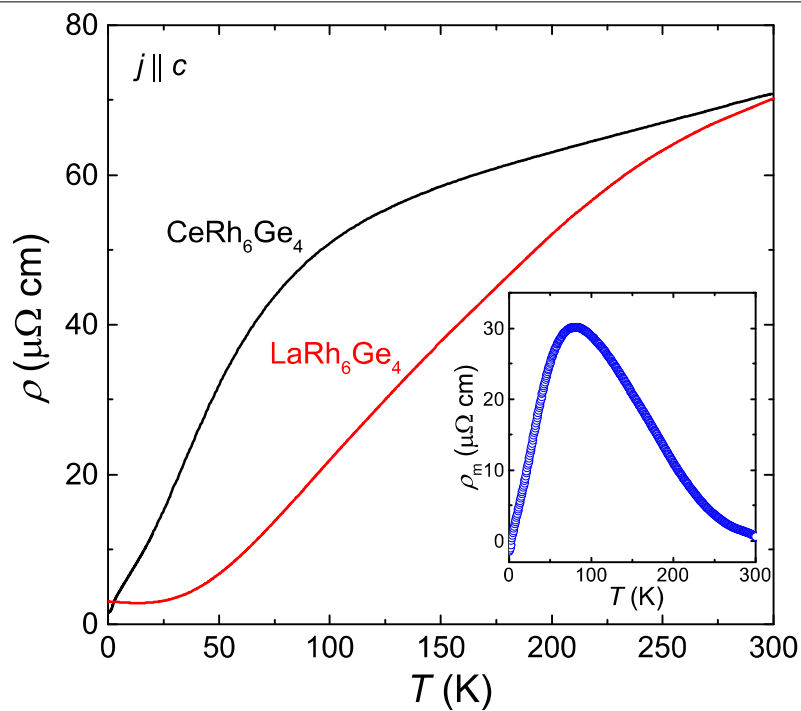
Correspondence and requests for materials should be addressed to M.S., P.C. or H.Y.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



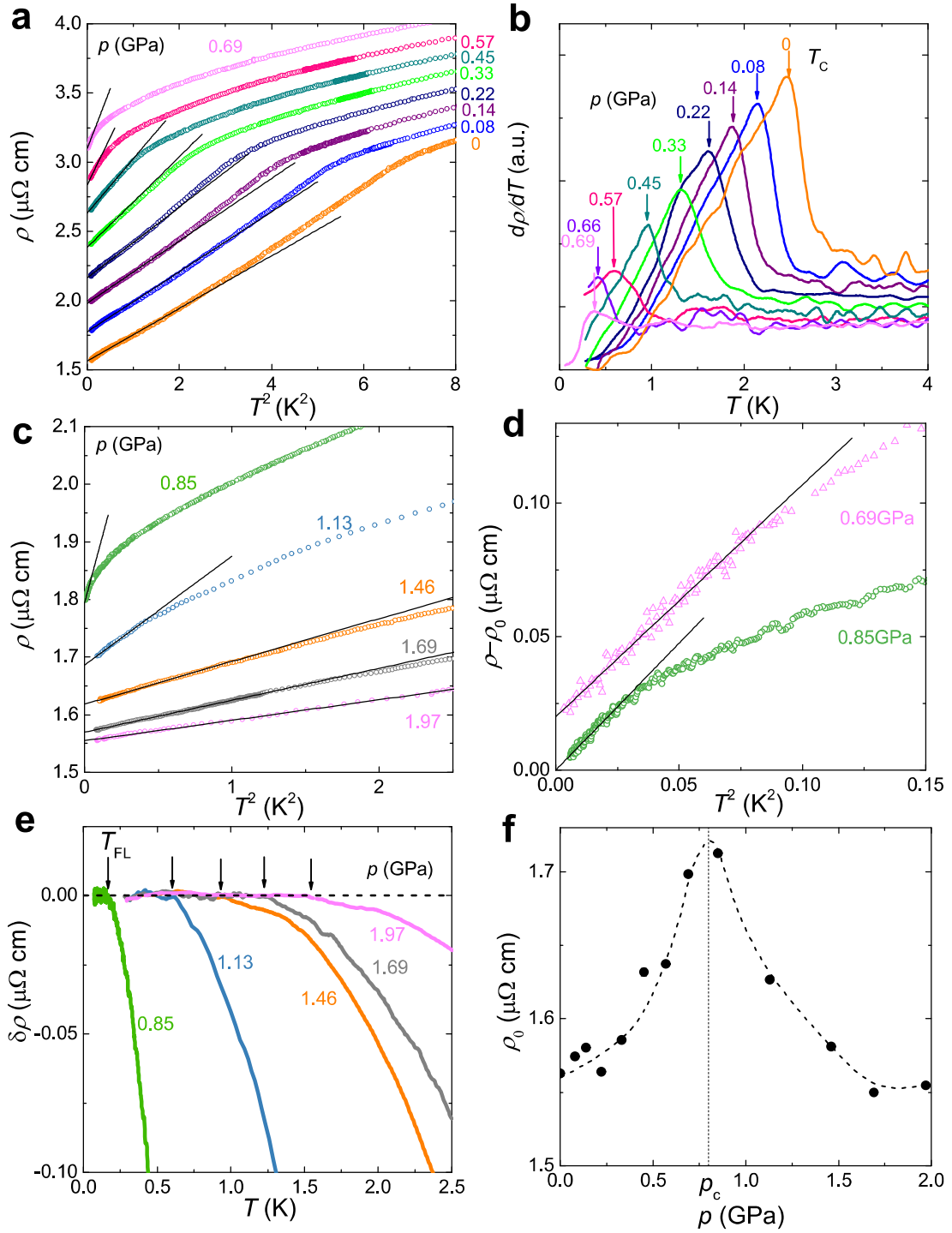
Extended Data Fig. 1 | Magnetic susceptibility and field-dependent magnetization. **a**, Temperature dependence of the magnetic susceptibility ($\chi(T)$) of CeRh₆Ge₄ in a field of 0.1 T applied both along the c axis and in the a - b plane, where both axes are plotted on a logarithmic scale. $\chi(T)$ is anisotropic across the whole temperature range; the a - b plane corresponds to the easy

direction. **b**, Magnetization loops measured at 3 K and 0.44 K, above and below T_c , respectively. In the FM state, the magnetization increases rapidly at low fields, reaching a value of around $0.28 \mu_B$ per Ce atom, which probably corresponds to the ordered moment, whereas at higher fields the magnetization increases more slowly.



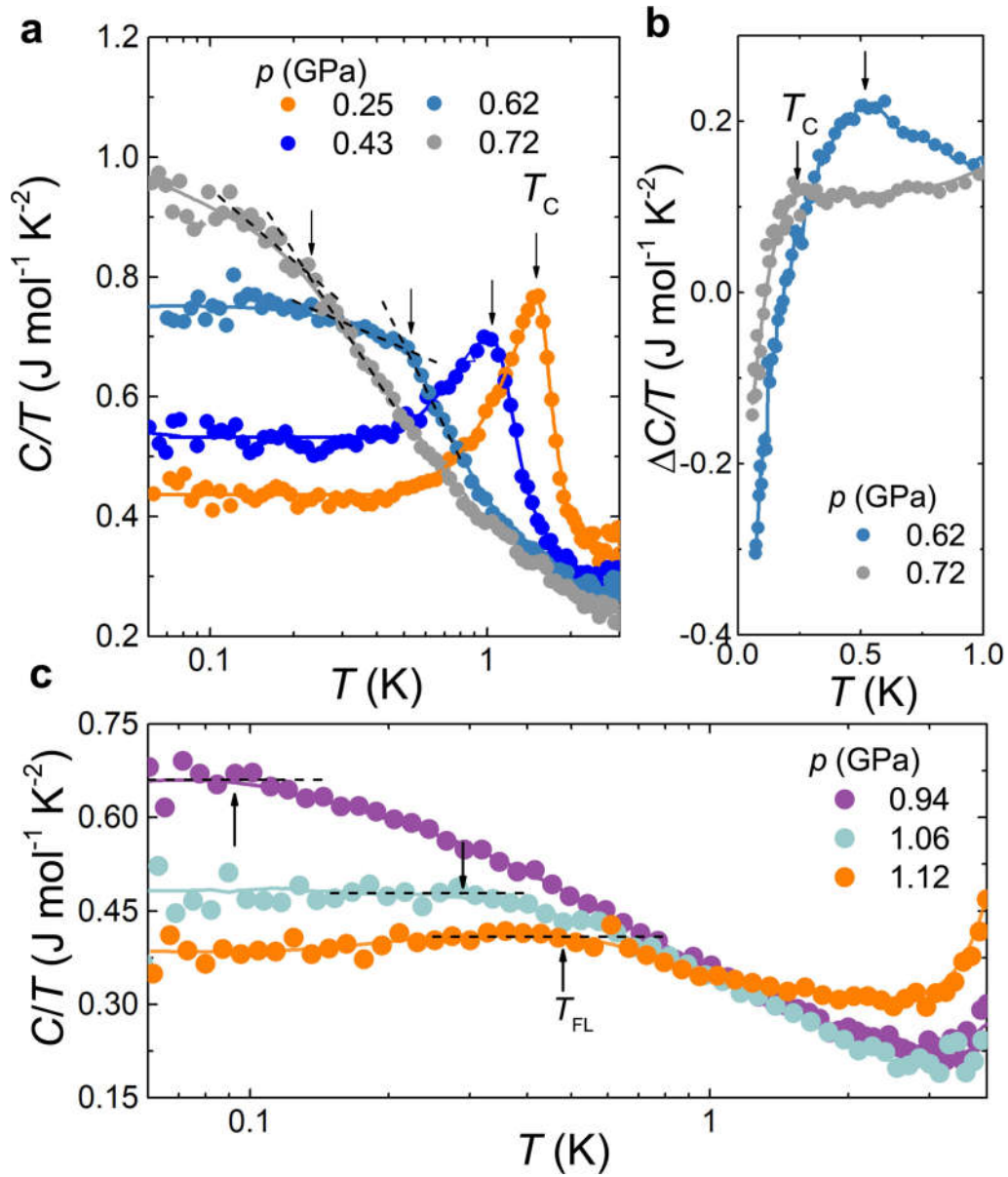
Extended Data Fig. 2 | Temperature-dependent resistivity at ambient pressure. Temperature dependence of the resistivity ($\rho(T)$) of CeRh_6Ge_4 and for the non-magnetic analogue LaRh_6Ge_4 , with the current along the c axis. The inset shows the magnetic contribution to the resistivity of CeRh_6Ge_4 (ρ_m),

obtained from subtracting the data of LaRh_6Ge_4 . This exhibits a broad maximum at around 80 K, probably as a consequence of both the crystalline electric field and Kondo effects.



Extended Data Fig. 3 | Analysis of the resistivity under pressure. **a**, Low-temperature $\rho(T)$ of CeRh_6Ge_4 , versus T^2 under pressures up to 0.69 GPa. For clarity, the data at consecutive pressures are offset vertically by $0.2 \mu\Omega \text{ cm}$. The low-temperature data in the magnetic state was fitted with a quadratic temperature dependence, $\rho(T) = \rho_0 + AT^2$, as shown by the solid black lines. **b**, The corresponding derivative $d\rho(T)/dT$, where the position of T_c was determined at each pressure from the position of the maximum, as indicated by the vertical arrows. a.u., arbitrary units. **c**, Low-temperature $\rho(T)$ versus T^2 of CeRh_6Ge_4 , at pressures above the QCP; the data at consecutive pressures are offset vertically by $0.02 \mu\Omega \text{ cm}$. The solid lines show the quadratic temperature dependence, indicating the occurrence of Fermi-liquid behaviour at low

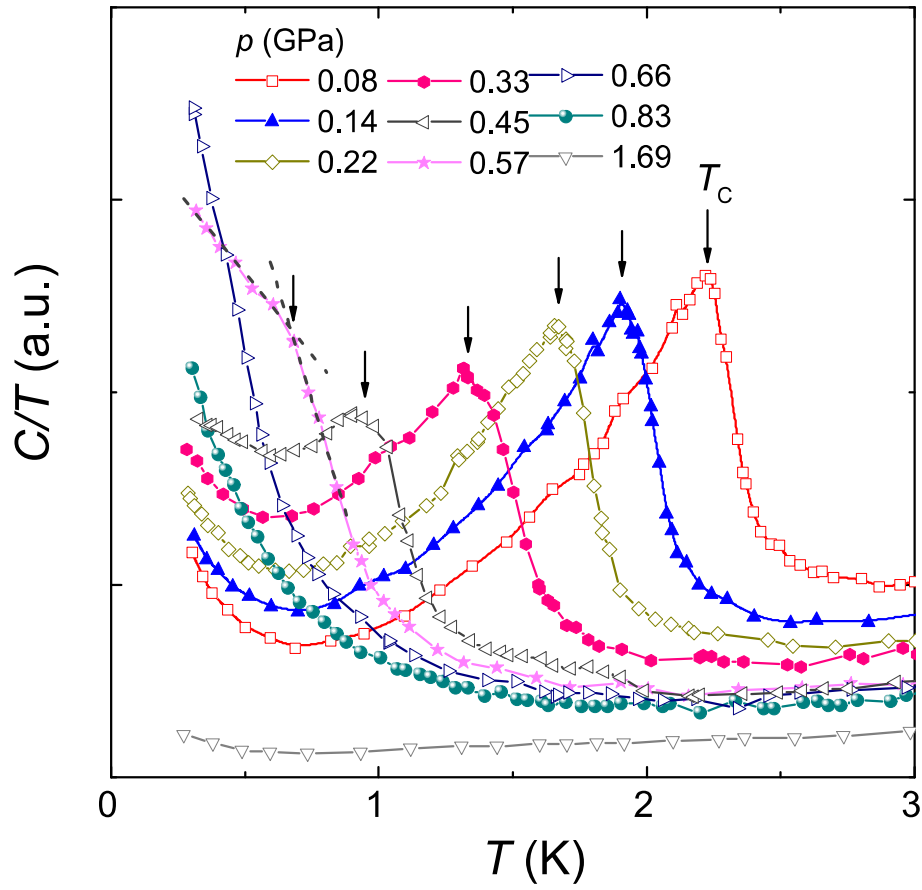
temperatures. **d**, Low-temperature enlargement of $\rho(T) - \rho_0$ for two pressures either side of the QCP, where the data at 0.69 GPa are vertically offset by $0.02 \mu\Omega \text{ cm}$. **e**, Resistivity as a function of temperature plotted as $\delta\rho = \rho - \rho_{\text{FL}}$, for various pressures p . ρ_{FL} is the Fermi-liquid contribution to the resistivity, obtained from fitting the low-temperature $\rho(T)$ with a quadratic temperature dependence. The deviation of $\delta\rho$ from zero indicates the onset of non-Fermi-liquid behaviour, and hence corresponds to T_{FL} , as marked by the vertical arrows. **f**, Pressure dependence of the residual resistivity ρ_0 , obtained from analysing the low-temperature $\rho(T)$ at various pressures, and where the error bars are smaller than the symbol size. This quantity reaches a maximum around the QCP.



Extended Data Fig. 4 | Analysis of the heat capacity under pressure.

a, Temperature dependence of the absolute value of the heat capacity as C/T , at various pressures below p_c . For pressures up to 0.72 GPa, T_c can be detected, as marked by the vertical arrows. At lower pressures this is determined from the peak positions, whereas close to p_c it is determined by the intersection of the solid lines indicated in the figure. **b**, The data for two pressures near p_c , after subtracting the data taken at 0.8 GPa to remove the logarithmic contribution to C/T . In both cases, the peak position of $\Delta C/T$ is in good agreement with the

value of T_c obtained from **a**. **c**, Low-temperature $C(T)/T$ for three pressures above the QCP. The strong increase with decreasing temperature corresponds to non-Fermi-liquid behaviour, whereas the flattening of $C(T)/T$ at low temperatures corresponds to the onset of Fermi-liquid behaviour. The position of the temperature below which Fermi-liquid behaviour occurs, T_{FL} , is highlighted by the vertical arrows, and is determined from the deviation from the near-temperature-independent behaviour marked by the dashed lines.



Extended Data Fig. 5 | The a.c. heat capacity under pressure. The a.c. heat capacity as C/T at various pressures up to 1.69 GPa. For pressures below 0.83 GPa, the position of T_c is marked by the vertical arrows. The dashed lines show the construction used to determine T_c near p_c . At 0.83 GPa, no transition

is detected down to the lowest measured temperature, 0.3 K; instead, C/T continues to increase with decreasing temperature. At 1.69 GPa, well above the QCP, C/T shows little temperature dependence. a.u., arbitrary units.

Tunable correlated Chern insulator and ferromagnetism in a moiré superlattice

<https://doi.org/10.1038/s41586-020-2049-7>

Received: 16 May 2019

Accepted: 11 December 2019

Published online: 4 March 2020

 Check for updates

Guorui Chen^{1,2}, Aaron L. Sharpe^{3,4}, Eli J. Fox^{4,5}, Ya-Hui Zhang⁶, Shaoxin Wang², Lili Jiang², Bosai Lyu^{7,8}, Hongyuan Li^{1,2}, Kenji Watanabe⁹, Takashi Taniguchi⁹, Zhiwen Shi^{7,8}, T. Senthil⁶, David Goldhaber-Gordon^{4,5}✉, Yuanbo Zhang^{8,10,11}✉ & Feng Wang^{1,2,12}✉

Studies of two-dimensional electron systems in a strong magnetic field revealed the quantum Hall effect¹, a topological state of matter featuring a finite Chern number C and chiral edge states^{2,3}. Haldane⁴ later theorized that Chern insulators with integer quantum Hall effects could appear in lattice models with complex hopping parameters even at zero magnetic field. The ABC-trilayer graphene/hexagonal boron nitride (ABC-TLG/hBN) moiré superlattice provides an attractive platform with which to explore Chern insulators because it features nearly flat moiré minibands with a valley-dependent, electrically tunable Chern number^{5,6}. Here we report the experimental observation of a correlated Chern insulator in an ABC-TLG/hBN moiré superlattice. We show that reversing the direction of the applied vertical electric field switches the moiré minibands of ABC-TLG/hBN between zero and finite Chern numbers, as revealed by large changes in magneto-transport behaviour. For topological hole minibands tuned to have a finite Chern number, we focus on quarter filling, corresponding to one hole per moiré unit cell. The Hall resistance is well quantized at $h/2e^2$ (where h is Planck's constant and e is the charge on the electron), which implies $C = 2$, for a magnetic field exceeding 0.4 tesla. The correlated Chern insulator is ferromagnetic, exhibiting substantial magnetic hysteresis and a large anomalous Hall signal at zero magnetic field. Our discovery of a $C = 2$ Chern insulator at zero magnetic field should open up opportunities for discovering correlated topological states, possibly with topological excitations⁷, in nearly flat and topologically nontrivial moiré minibands.

Moiré superlattices in van der Waals heterostructures have emerged as a powerful tool for engineering quantum phenomena, because the periodic moiré potential defines new length and energy scales^{8–10}. Notably, nearly flat electronic bands can be realized in different moiré superlattice systems, which offer exciting opportunities to realize a wide variety of correlation physics^{5,6}. For example, correlated insulators and superconductivity have been reported in magic-angle twisted bilayer graphene^{11–13} and in ABC-TLG/hBN moiré superlattices^{14,15}, and spontaneous ferromagnetism and an anomalous Hall effect, apparently corresponding to an incipient Chern insulator, have been observed in twisted bilayer graphene with an aligned hBN layer¹⁶. Recent theories suggest that correlated topological phenomena could emerge in such graphene moiré superlattices, where a non-trivial band topology coexists with the nearly flat moiré miniband^{5,6,17,18}. Pristine ABC-TLG, because of its cubic band and therefore a rather flat dispersion at low energy, can already exhibit strong correlations¹⁹ and can potentially

host spontaneous quantum Hall states²⁰. The moiré superlattice in ABC-TLG/hBN heterostructure further creates isolated flat moiré minibands, which enhances the electron–electron correlation and topological effects in the system. This ABC-TLG/hBN heterostructure provides a particularly attractive platform with which to explore correlated topological phenomena because not only the electron density but also the bandwidth and topology of the moiré minibands can be conveniently controlled by electrostatic gating^{5,6,14}.

Here we report experimental observation of a correlated Chern insulator and ferromagnetism in ABC-TLG/hBN. Upon tuning the vertical displacement field, we show that the magneto-transport in an ABC-TLG/hBN moiré superlattice exhibits distinct behaviours for trivial minibands ($C = 0$) compared to topological minibands ($C \neq 0$). A correlated Chern insulator with $C = 2$ quantum anomalous Hall effect²¹ emerges around 1/4 filling of the topological hole miniband when the bandwidth is sufficiently narrowed by applying a displacement field

¹Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ²Department of Physics, University of California at Berkeley, Berkeley, CA, USA. ³Department of Applied Physics, Stanford University, Stanford, CA, USA. ⁴Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. ⁵Department of Physics, Stanford University, Stanford, CA, USA. ⁶Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁷Key Laboratory of Artificial Structures and Quantum Control (Ministry of Education), School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China. ⁸Collaborative Innovation Center of Advanced Microstructures, Nanjing, China. ⁹National Institute for Materials Science, Tsukuba, Japan. ¹⁰State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai, China. ¹¹Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai, China. ¹²Kavli Energy NanoSciences Institute at the University of California, Berkeley, and the Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ✉e-mail: goldhaber-gordon@stanford.edu; zhyb@fudan.edu.cn; fengwang76@berkeley.edu

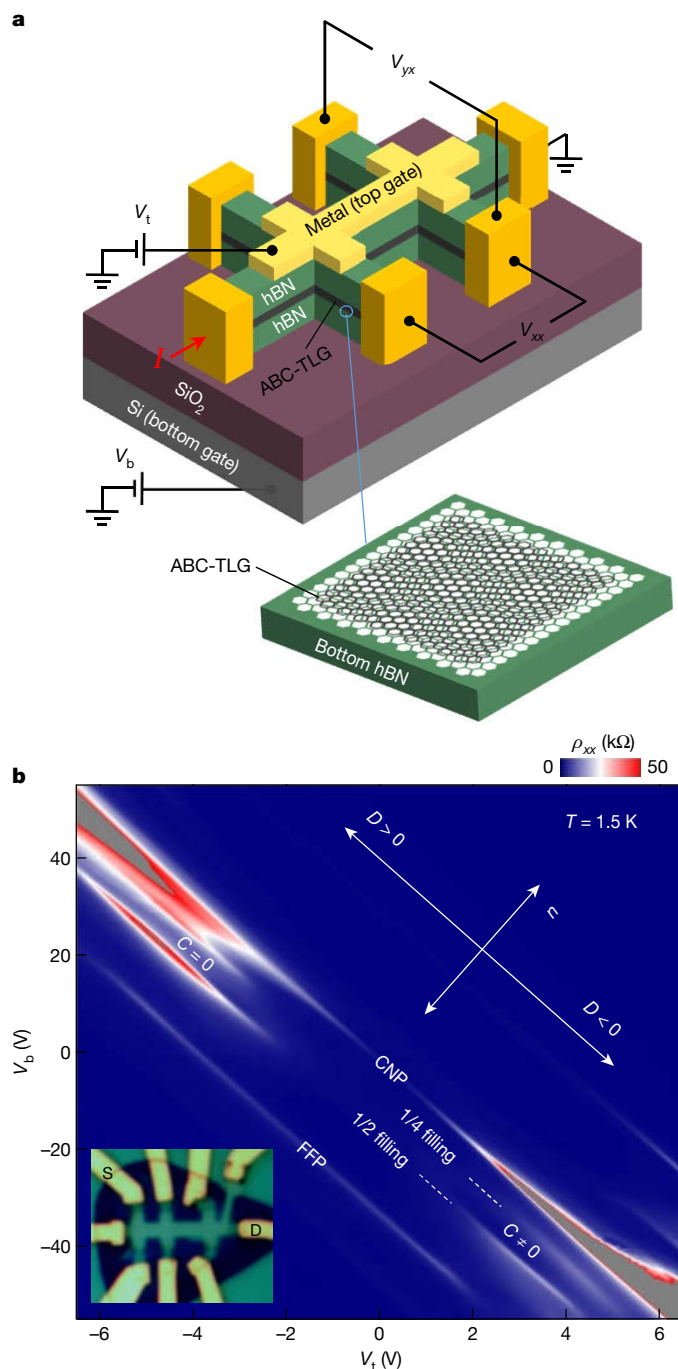


Fig. 1 | ABC-TLG/hBN moiré superlattice and tunable Chern bands.

a, Schematic of the dual-gated ABC-TLG/hBN moiré superlattice Hall bar device and measurement configuration. The inset shows that the moiré pattern exists between ABC-TLG and bottom hBN. **b**, Colour plot of the longitudinal resistivity ρ_{xx} as a function of V_t and V_b at $T = 1.5$ K. The arrows show the direction of changing doping n and displacement field D , respectively. In addition to the band insulating states (characterized by the resistance peaks) at the charge neutral point (CNP) and fully filled point (FFP), tunable correlated insulator states also emerge at 1/4 filling and 1/2 filling of the hole minibands at large displacement field $|D|$. It has been predicted theoretically^{5,6} that the hole miniband is topological (that is, Chern number $C \neq 0$) for $D < 0$ and trivial ($C = 0$) for $D > 0$. The inset shows the optical image of the device.

in one direction. The correlated Chern insulator spontaneously breaks time-reversal symmetry, exhibiting strong ferromagnetic hysteresis and a zero-field anomalous Hall resistance over 8 k Ω . The experimentally observed $C = 2$ Chern band can be understood theoretically by

incorporating electron–electron interaction effects on the quasi-particle band structure of ABC-TLG/hBN moiré minibands.

We fabricated two ABC-TLG/hBN moiré superlattice devices following the method described in ref. ¹⁴. In brief, the ABC-TLG domain is identified by scanning near-field infrared nanoscopy²² (Extended Data Fig. 1), and is isolated from adjacent ABA domains by atomic force microscope cutting²³. The isolated ABC-TLG is then encapsulated in exfoliated hBN crystals, where one hBN crystal is aligned with the ABC-TLG to form the moiré superlattice. The ABC-TLG/hBN heterostructures are fabricated into a Hall bar geometry with one-dimensional edge contacts, a metal top gate, and a degenerately doped silicon bottom gate following standard nanofabrication procedures²⁴. A schematic image and an optical image of device I (data shown in the main text is from device I) are shown in the insets to Fig. 1a, b (and Extended Data Fig. 2). Gate voltages V_t and V_b are applied to the metal top gate and the Si bottom gate, respectively. The dual-gate configuration allows us to independently control the doping and the miniband bandwidth of the ABC-TLG/hBN heterostructure^{25–27}: the doping relative to the charge neutrality point is set by $n = (D_b - D_t)/e$, and the miniband bandwidth is tuned by the applied vertical displacement field $D = (D_b + D_t)/2$. Here $D_b = +\epsilon_b(V_b - V_b^0)/d_b$ and $D_t = -\epsilon_t(V_t - V_t^0)/d_t$ are the vertical displacement fields below and above the ABC-TLG/hBN moiré superlattice, respectively, $\epsilon_{b(t)}$ and $d_{b(t)}$ are the dielectric constant and thickness of the bottom (top) dielectric layers, and $V_{b(t)}^0$ is the effective offset in the bottom (top) gate voltages caused by environment-induced carrier doping. The longitudinal resistivity ρ_{xx} is obtained by $\rho_{xx} = (W/L)V_{xx}/I$, where $W = 1 \mu\text{m}$ is the channel width and $L = 4 \mu\text{m}$ is the channel length, and the Hall resistivity ρ_{yx} is obtained by $\rho_{yx} = V_{yx}/I$ (the measurement configuration for the longitudinal and Hall voltages, V_{xx} and V_{yx} , respectively, is shown in Fig. 1a).

As the voltages applied to the gates are tuned, measurements of ρ_{xx} reveal several resistance peaks (Fig. 1b) in the ABC-TLG/hBN device across the parameter space controlled by V_t and V_b . In addition to the peaks corresponding to band insulating states at the charge neutrality point and fully filled point, tunable correlated insulator states emerge at 1/4 filling and 1/2 filling of the hole miniband (that is, one and two holes per moiré unit cell) when a finite displacement field $|D|$ narrows the moiré minibands. There are two apparent asymmetries of the correlated insulator states in ABC-TLG/hBN at 1/4 and 1/2 charge fillings: between the electron and hole minibands, and between positive and negative D . Prominent correlated insulator states are observed in the hole minibands but not in the electron minibands because the hole miniband has a much smaller bandwidth for finite $|D|$ (ref. ¹⁴). The asymmetry between the positive and negative D fields arises from the fact that the moiré superlattice exists only between the ABC-TLG and the bottom hBN in this device (Fig. 1a). Interestingly, the direction of the displacement field has been predicted to determine not only the relative bandwidth but also the topology of the hole miniband. For a device with moiré superlattice between ABC-TLG and the bottom (top) hBN, a positive (negative) D leads to a trivial hole miniband with $C = 0$ and smaller bandwidth, while a negative (positive) D leads to a topological hole miniband with $C \neq 0$ and larger bandwidth^{5,6}. The difference in bandwidth and topology is reflected in the electrical transport behaviour: in the device with the moiré superlattice at the bottom hBN, we observe stronger correlated insulators in the trivial hole miniband with positive D , because holes are easier to localize in the narrower trivial band than in the broader topological band with negative D (ref. ¹⁷). Our previous studies have shown that superconductivity can emerge when we dope the 1/4 filling correlated insulator state in selected parameter spaces of the trivial hole miniband¹⁵.

To better probe the topological aspects of the moiré minibands, we turn to magneto-transport studies. At $D = 0$, the correlation effect in the system is relatively weak, and the magneto-transport data exhibit well defined quantum Hall states and a Landau fan diagram at low magnetic fields (see Extended Data Fig. 4), demonstrating the very

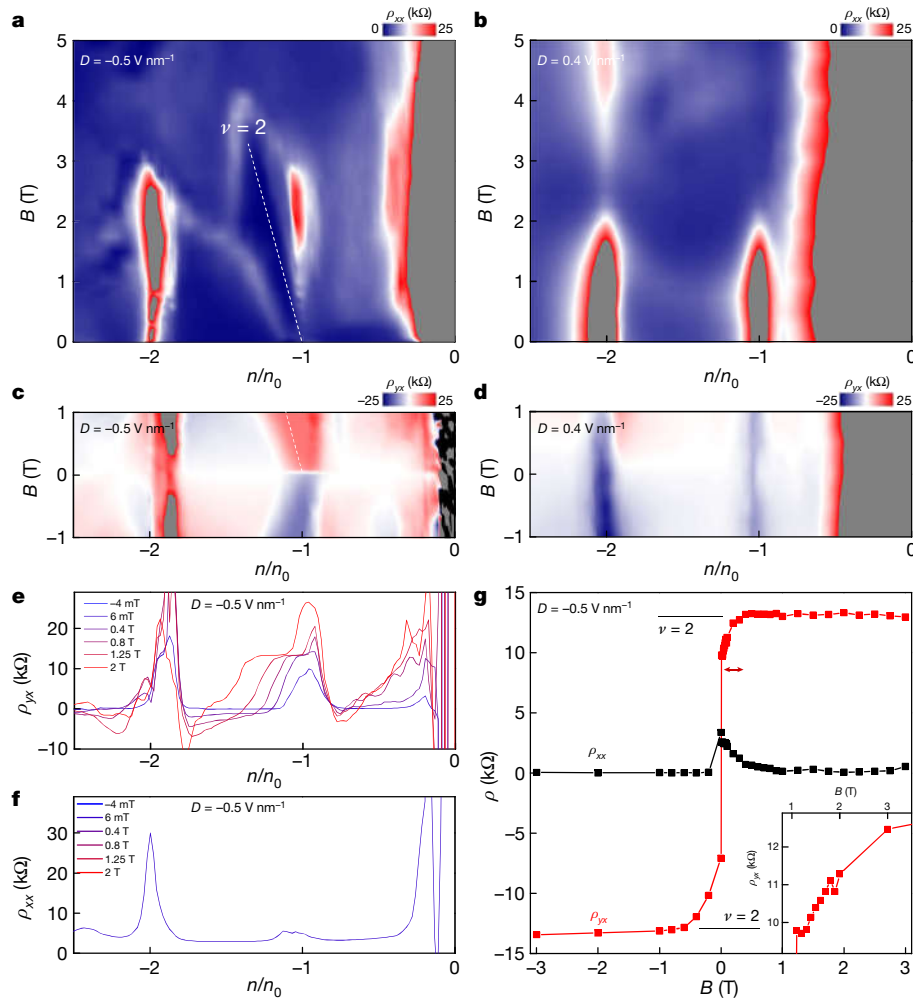


Fig. 2 | Quantum Hall effect from the correlated $C=2$ Chern insulator.

a, c, Colour plot of ρ_{xx} (**a**) and ρ_{yx} (**c**) as a function of carrier density and magnetic field for the topological hole miniband at $D = -0.5$ V nm $^{-1}$ and $T = 0.06$ K. The experimental data at $T = 1.5$ K are qualitatively similar. **b, d,** Corresponding ρ_{xx} (**b**) and ρ_{yx} (**d**) plots for the trivial hole miniband at $D = 0.4$ V nm $^{-1}$ and $T = 1.5$ K. n_0 corresponds to the carrier density of the 1/4 filling of the first miniband. No quantum Hall signatures are present in the trivial hole miniband, whereas a $\nu = 2$ quantum Hall effect characterized by a minimum of

high quality of our sample. With large displacement field, the moiré miniband bandwidth becomes narrower, and the dominant electron–electron interaction dramatically changes the magneto-transport behaviour. Figure 2a, b displays the colour plot of ρ_{xx} as a function of the hole doping and the vertical magnetic field B for $D = -0.5$ V nm $^{-1}$ (at $T = 0.06$ K) and $D = 0.4$ V nm $^{-1}$ (at $T = 1.5$ K), respectively. Figure 2c, d shows the corresponding Hall resistivity ρ_{yx} data. Experimental data for $D = -0.5$ V nm $^{-1}$ at 1.5 K exhibit qualitatively similar behaviour to those at 0.06 K (see Extended Data Fig. 3). We have used $n_0 = 5.25 \times 10^{11}$ cm $^{-2}$ as the unit of carrier density, which corresponds to one hole per moiré lattice site (that is, 1/4 filling). The magneto-transport data exhibit distinct behaviours for the topological moiré miniband at negative D and the trivial miniband at positive D (refs. ^{5,6}). Specifically, a strong quantum Hall state emerges from the 1/4 filling point for $D = -0.5$ V nm $^{-1}$ but not for $D = 0.4$ V nm $^{-1}$. The dashed line in Fig. 2a traces the minimum in ρ_{xx} following the relation $n = \nu eB/h$ for $\nu = 2$. This quantum Hall state is well developed at very low magnetic fields, and originates from the 1/4 filling resistive state at zero magnetic field (Fig. 2a). At the same time, ρ_{yx} is very large at weak magnetic fields and exhibits a jump in value when the magnetic field switches sign across $B = 0$ T (Fig. 2c). By contrast,

ρ_{xx} and a quantized ρ_{yx} emerges from 1/4 filling of the topological hole miniband. **e, f,** Horizontal line cuts of **a** and **c**, respectively. **e** shows that ρ_{yx} is well quantized beyond $B = 0.4$ T. An offset of 2.5 k Ω is applied for each stack in **f**. **g,** Line cut of ρ_{xx} and ρ_{yx} along the quantum Hall state (denoted by the dashed lines in **a** and **e**) shows that ρ_{yx} reaches a quantized value of $\nu = 2$ at 0.4 T, and a large ρ_{xx} persists down to zero field. It represents a quantum anomalous Hall state for the $C=2$ correlated Chern insulator at 1/4 filling. The inset shows a zoomed-in plot of ρ_{yx} at small magnetic field.

stronger correlated insulator states are observed for $D = 0.4$ V nm $^{-1}$, but no signatures of quantum oscillations or quantum Hall effects are present (Fig. 2b). In addition, Fig. 2d shows that the Hall resistivity signal tends to be rather small for all hole doping at $D = 0.4$ V nm $^{-1}$. (The relatively large ρ_{yx} signals at 1/4 and 1/2 fillings are artefacts caused by crosstalk from the large ρ_{xx} of the correlated insulator states, and they do not change sign when the magnetic field is reversed.)

Figure 2e, f shows ρ_{yx} and ρ_{xx} as a function of density for a few representative magnetic field values, corresponding to horizontal line cuts in Fig. 2c and a, respectively. ρ_{yx} is well quantized for magnetic field larger than 0.4 T at the value of 13.0 ± 0.2 k Ω , that is, the expected quantized value of $h/2e^2 = 12.9$ k Ω is within the empirical uncertainty. ρ_{xx} exhibits a corresponding minimum in the quantum Hall state, with a minimum resistivity less than 60 Ω at 2 T. Figure 2g further displays ρ_{yx} and ρ_{xx} as a function of the magnetic field along the quantum Hall state following the dashed line in Fig. 2c, with the inset showing a zoomed-in plot of ρ_{yx} between 0 to 0.2 T. ρ_{yx} smoothly reaches the quantized value at 0.4 T. ρ_{yx} maintains a large though not quantized value all the way to zero magnetic field, and a large jump of ρ_{yx} is observed when the magnetic field changes sign.

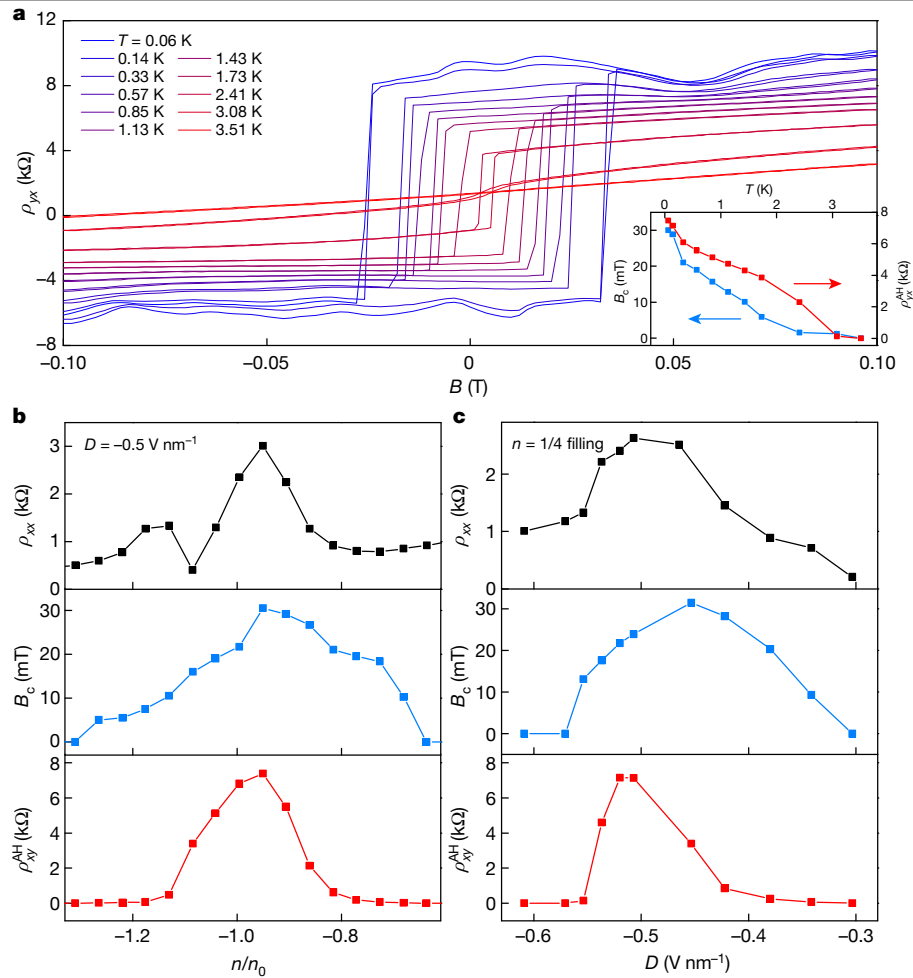


Fig. 3 | Anomalous Hall effect and ferromagnetism. a, Magnetic-field-dependent ρ_{yx} at 1/4 filling and $D = -0.5 \text{ V nm}^{-1}$ at different temperatures. The Hall resistivity displays a clear anomalous Hall signal with strong ferromagnetic hysteresis. At the base temperature of $T = 0.06 \text{ K}$, the anomalous Hall signal can be as high as $\rho_{yx}^{\text{AH}} = 8 \text{ k}\Omega$ and the coercive field is $B_c = 30 \text{ mT}$. The inset shows the extracted coercive field B_c and anomalous Hall signal ρ_{yx}^{AH} as a

function of temperature. **b**, The evolution of ρ_{xx} , B_c and ρ_{yx}^{AH} as a function of hole doping at $D = -0.5 \text{ V nm}^{-1}$, $T = 0.06 \text{ K}$. The strongest anomalous Hall signal is observed close to $n = n_0$. **c**, The evolution of ρ_{xx} , B_c and ρ_{yx}^{AH} as a function of the displacement field D at $n = n_0$, $T = 0.06 \text{ K}$. The strongest anomalous Hall signal is observed when the device is most insulating (that is, largest ρ_{xx}).

The $\nu = 2$ quantum Hall state at 1/4 filling at $D = -0.5 \text{ V nm}^{-1}$ cannot be explained by a conventional integer quantum Hall effect from single-particle Landau levels. Instead, we argue that it represents a quantum anomalous Hall state from a correlated Chern insulator. First, this quantum Hall state only exists at negative D , where the miniband is predicted to have a non-trivial Chern number, and is absent at the positive D , where the band is predicted to be trivial. Second, it is well established that the lowest single-particle Landau level in ABC-TLG should be a $\nu = 3$ state owing to a winding number of 3 close to the valence band maximum²⁸. Third, only one quantum Hall state is observed anywhere, and the quantized Hall resistivity appears to start at very low magnetic field. If the observed quantum Hall state of 1/4 filling at $D = -0.5 \text{ V nm}^{-1}$ is from the lowest single-particle Landau level, similar Landau levels should also exist close to the charge neutrality point and the 1/2 filling correlated insulators, and higher Landau levels should be observable. (See Extended Data Fig. 4 for a single-particle Landau fan diagram in the same device, where $D = 0$ and the electron correlation is weak.) Finally, an apparent non-zero quantum Hall-like gap was observed for the 1/4 filling Chern insulator state down to $B = 0 \text{ T}$; the size of the gap continuously increases with increasing B (see Methods and Extended Data Fig. 5). All our data can be naturally explained by a $\nu = 2$ Chern insulator state at 1/4 filling. Such a $C = 2$ correlated Chern insulator should feature quantized Hall resistivity ρ_{yx} and a corresponding magnetic

field dependent carrier density based on the Streda formula²⁹. This Chern insulator at 1/4 filling is a strongly correlated state that breaks the valley degeneracy and fills only the $C = 2$ electronic band in one valley. The nearly flat and tunable moiré minibands in the ABC-TLG/hBN moiré heterostructure are critical for the realization of such a correlated topological state.

The correlated Chern insulator, persisting to zero magnetic field, spontaneously breaks the time-reversal symmetry and can generate valley-flavour ferromagnetism at 1/4 filling. Indeed, ferromagnetism and strong anomalous Hall signals emerge from the Chern insulator state at zero magnetic field. Figure 3a shows the temperature-dependent Hall resistivity when a small perpendicular B is swept between -0.1 T and 0.1 T . The Hall resistivity displays a clear anomalous Hall signal with strong ferromagnetic hysteresis. At $B = 0 \text{ T}$, ρ_{yx} is non-zero and depends on the magnetic field sweep direction, a defining ferromagnetic feature. At the base temperature of $T = 0.06 \text{ K}$, the anomalous Hall signal reaches a maximum of $\rho_{yx}^{\text{AH}} = 8 \text{ k}\Omega$ and a coercive field as large as $B_c = 30 \text{ mT}$. The inset in Fig. 3a shows the temperature dependence of ρ_{yx}^{AH} and B_c : both signals decrease monotonically with increasing temperature, reaching zero at $T = 3.5 \text{ K}$. The zero magnetic field ρ_{yx}^{AH} is already close to $12.9 \text{ k}\Omega$. An almost perfect quantization of the $C = 2$ quantum anomalous Hall¹⁸ Chern insulator appears at a magnetic field as low as 0.4 T .

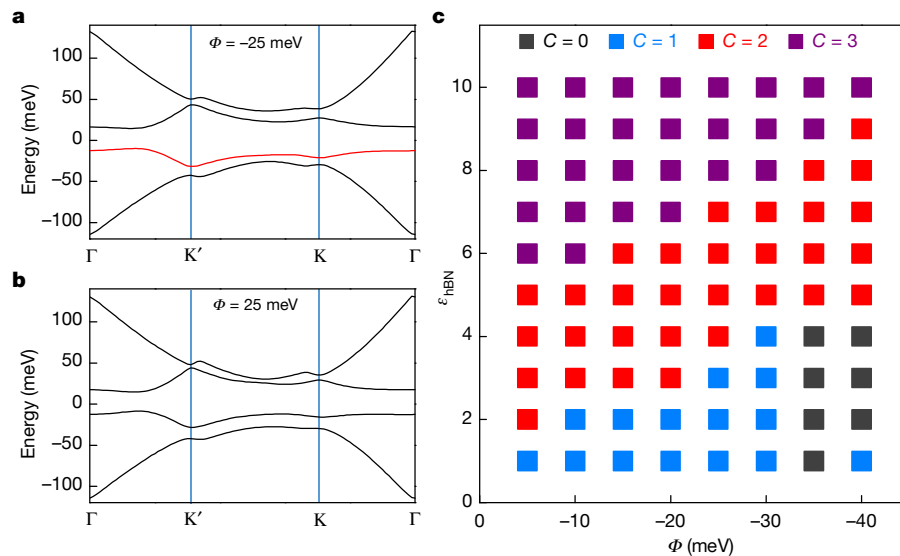


Fig. 4 | Calculated Chern number including the electron–electron interaction effects. **a, b**, Calculated single-particle band structure of the ABC-TLG/hBN moiré superlattice for $\Phi = -25$ meV and 25 meV, respectively. Here Φ is the energy difference between the top and bottom layers of ABC-TLG, and $\Phi = -25$ meV corresponds to the vertical displacement field around $D = -0.5$ V nm $^{-1}$. The red line highlights the topological hole miniband for

The ferromagnetism is tunable by n and D and appears only in a limited parameter space of n and D . In Fig. 2e, ρ_{yx} near $n = n_0$ presents different signs at $B = -4$ mT and 6 mT, which is much smaller than B_c . For this measurement, the magnetic field is fixed and carrier density is swept from a non-ferromagnetic state to a ferromagnetic state, which leads to ρ_{yx} with different signs even in small positive and negative magnetic fields. A clearer n -dependence of ρ_{xx} , B_c and ρ_{yx}^{AH} at $D = -0.5$ V nm $^{-1}$ at the base temperature is shown in Fig. 3b by sweeping the magnetic field at different fixed n . B_c and ρ_{yx}^{AH} both have maximum values close to $n = n_0$. However, ρ_{yx}^{AH} shows a stronger carrier doping dependence and decreases to almost zero at $n = n_0 \pm 0.2n_0$, while B_c decreases to zero at $n = n_0 \pm 0.35n_0$. ρ_{xx} displays an unusual behaviour with both a resistance peak and a resistance dip close to $n = n_0$, the origin of which requires further experimental and theoretical study. Figure 3c shows the D -dependence of ρ_{xx} , B_c and ρ_{yx}^{AH} at $n = n_0$. ρ_{xx} shows a maximum at $D = -0.5$ V nm $^{-1}$, which might be due to the narrowest bandwidth and strongest correlation effects at this displacement field¹⁴. ρ_{yx}^{AH} also shows a maximum at $D = -0.5$ V nm $^{-1}$, suggesting the importance of electron–electron correlations to the observed anomalous Hall signal. A finite ρ_{yx}^{AH} can be observed with D between -0.3 V nm $^{-1}$ and -0.57 V nm $^{-1}$. A non-zero B_c is present in the same D range, although the maximum B_c appears at $D = -0.45$ V nm $^{-1}$.

The observed $C = 2$ correlated Chern insulator can be understood theoretically from the topological moiré minibands when the electron–electron interactions are considered. Previous theoretical calculations predict a valley Chern number $C = 3$ for the single-particle hole miniband for negative D (refs.^{5,6}), but our results suggest that interaction effects can renormalize the valley Chern number. Figure 4a, b shows the single-particle band structures of the lowest few moiré minibands in ABC-TLG/hBN moiré superlattices for positive and negative displacement fields. For the negative D values (supporting a non-zero valley Chern number), the valence band overlaps with the remote lower band (see Fig. 4a). We incorporate the interaction effects in Hartree–Fock theory. When the valence band is close to the band below (at large $|D|$) or when the interaction strength is sufficiently strong (with small dielectric constant), the self-energy corrections mix the valence band and the lower band, leading to reduction of the Chern number to $C = 2$. As shown in Fig. 4c, when the dielectric constant is around 4 (effective screening from the

$\Phi = -25$ meV. **c**, Calculated Chern number of the hole miniband as a function of the energy difference Φ and the effective dielectric constant ϵ_{hBN} after including the electron–electron interaction effects using the Hartree–Fock approximation. The resulting band Chern number can be 2 for parameters close to the experimental device where $\Phi \approx -25$ meV and $\epsilon_{hBN} \approx 4$.

dielectric constant of hBN), the valley Chern number is expected to be 2 for a large range of displacement field values.

Our observation of a tunable $C = 2$ Chern insulator in ABC-TLG/hBN moiré superlattice provides an opportunity to explore correlated topological states in van der Waals moiré heterostructures. For example, fractional Chern insulators and non-Abelian states could emerge from strong correlations in nearly flat topological minibands once the quality of moiré heterostructures is further improved. In particular, the flat $C = 2$ Chern band has the potential to host novel fractional Chern insulator states beyond the fractional quantum Hall paradigm^{30,31}.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2049-7>.

1. Klitzing, K. v., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
2. Thouless, D. J., Kohmoto, M., Nightingale, M. P. & den Nijs, M. Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
3. Xiao, D., Chang, M.-C. & Niu, Q. Berry phase effects on electronic properties. *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
4. Haldane, F. D. M. Model for a quantum Hall effect without Landau levels: condensed-matter realization of the ‘parity anomaly’. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
5. Chittari, B. L., Chen, G., Zhang, Y., Wang, F. & Jung, J. Gate-tunable topological flat bands in trilayer graphene boron-nitride moiré superlattices. *Phys. Rev. Lett.* **122**, 016401 (2019).
6. Zhang, Y.-H., Mao, D., Cao, Y., Jarillo-Herrero, P. & Senthil, T. Nearly flat Chern bands in moiré superlattices. *Phys. Rev. B* **99**, 075127 (2019).
7. Bergholtz, E. J. & Liu, Z. Topological flat band models and fractional Chern insulators. *Int. J. Mod. Phys. B* **27**, 1330017 (2013).
8. Dean, C. R. et al. Hofstadter’s butterfly and the fractal quantum Hall effect in moiré superlattices. *Nature* **497**, 598–602 (2013).
9. Ponomarenko, L. A. et al. Cloning of Dirac fermions in graphene superlattices. *Nature* **497**, 594–597 (2013).
10. Hunt, B. et al. Massive Dirac fermions and Hofstadter butterfly in a van der Waals heterostructure. *Science* **340**, 1427–1430 (2013).
11. Cao, Y. et al. Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature* **556**, 80–84 (2018).

12. Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).
13. Lu, X. et al. Superconductors, orbital magnets, and correlated states in magic angle bilayer graphene. *Nature* **574**, 653–657 (2019).
14. Chen, G. et al. Evidence of a gate-tunable Mott insulator in a trilayer graphene moiré superlattice. *Nat. Phys.* **15**, 237 (2019).
15. Chen, G. et al. Signatures of tunable superconductivity in a trilayer graphene moiré superlattice. *Nature* **572**, 215–219 (2019).
16. Sharpe, A. L. et al. Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene. *Science* **365**, 605–608 (2019).
17. Zhang, Y.-H. & Senthil, T. Bridging Hubbard model physics and quantum Hall physics in trilayer graphene hBN moiré superlattice. *Phys. Rev. B* **99**, 205150 (2019).
18. Liu, J., Ma, Z., Gao, J. & Dai, X. Quantum valley Hall effect, orbital magnetism, and anomalous hall effect in twisted multilayer graphene systems. *Phys. Rev. X* **9**, 031021 (2019).
19. Lee, Y. et al. Competition between spontaneous symmetry breaking and single-particle gaps in trilayer graphene. *Nat. Commun.* **5**, 5656 (2014).
20. Zhang, F., Jung, J., Fiete, G. A., Niu, Q. & MacDonald, A. H. Spontaneous quantum Hall states in chirally stacked few-layer graphene systems. *Phys. Rev. Lett.* **106**, 156801 (2011).
21. Chang, C.-Z. et al. Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* **340**, 167–170 (2013).
22. Ju, L. et al. Topological valley transport at bilayer graphene domain walls. *Nature* **520**, 650–655 (2015).
23. Li, H. et al. Electrode-free anodic oxidation nanolithography of low-dimensional materials. *Nano Lett.* **18**, 8011–8015 (2018).
24. Wang, L. et al. One-dimensional electrical contact to a two-dimensional material. *Science* **342**, 614–617 (2013).
25. Zhang, Y. et al. Direct observation of a widely tunable bandgap in bilayer graphene. *Nature* **459**, 820–823 (2009).
26. Zou, K., Zhang, F., Clapp, C., MacDonald, A. H. & Zhu, J. Transport studies of dual-gated ABC and ABA trilayer graphene: band gap opening and band structure tuning in very large perpendicular electric fields. *Nano Lett.* **13**, 369–373 (2013).
27. Lui, C. H., Li, Z., Mak, K. F., Cappelluti, E. & Heinz, T. F. Observation of an electrically tunable band gap in trilayer graphene. *Nat. Phys.* **7**, 944–947 (2011).
28. Zhang, L., Zhang, Y., Camacho, J., Khodas, M. & Zaliznyak, I. The experimental observation of quantum Hall effect of $l=3$ chiral quasiparticles in trilayer graphene. *Nat. Phys.* **7**, 953–957 (2011).
29. MacDonald, A. The Quantum Hall Effects. In *Quantum Coherence in Mesoscopic Systems* (ed. Kramer, B.) 195–217 (Springer, 1991).
30. Liu, Z., Bergholtz, E. J., Fan, H. & Läuchli, A. M. Fractional Chern insulators in topological flat bands with higher Chern number. *Phys. Rev. Lett.* **109**, 186805 (2012).
31. Wang, Y.-F., Yao, H., Gong, C.-D. & Sheng, D. N. Fractional quantum Hall effect in topological flat bands with Chern number two. *Phys. Rev. B* **86**, 201101 (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Transport measurements

The ultralow temperature measurement is performed in a dilution refrigerator. Low-temperature electronic filtering, including microwave filters, low-pass resistor-capacitor filters, and thermal meanders, is used to anchor the electron temperature as well as to prevent quasiparticle excitations from high-frequency noise. Stanford Research Systems SR830 lock-in amplifiers with NF Corporation LI-75A voltage preamplifiers are used to measure the resistivity of the device with an alternating-current bias current of 0.5 nA at a frequency of 7 Hz.

The nature of the correlated insulator at 1/4 filling for the topological side

Because of the Wannier obstruction caused by the valley Chern number, a standard Mott insulator with localized charge is impossible¹⁷. Therefore, the physics on the topological side is essentially different from that on the trivial side, despite the similarity between the band structures. Because a narrow Chern band is analogous to a Landau level, the physics on the topological side is similar to that of quantum Hall systems with spin and valley degeneracies. At 1/4 filling, when the band is sufficiently flat, a single fully filled spin- and valley-polarized Chern band is favoured⁶, similar to the “quantum Hall ferromagnetism” found in Landau levels³². A valley-polarized Chern insulator matches the current transport experiment quite well. Ideally, a fully filled Chern band leads to quantized Hall conductivity $\sigma_{yx} = 2e^2/h$. At zero magnetic field, domains formed by the two degenerate valleys can cause a finite ρ_{xx} on the order of h/e^2 . Upon increasing the magnetic field to about 0.4 T, the valley Zeeman coupling¹⁷ can align the domains and lead to perfect quantization of the Hall conductivity.

Ferromagnetism and an anomalous Hall effect have also recently been observed in a near-magic-angle twisted bilayer graphene moiré superlattice at 3/4 filling of the conduction miniband, where an hBN cladding layer appears to be aligned with the proximate graphene¹⁶. These effects are likely to have a similar origin in both systems, where electron–electron interactions create a spontaneous valley polarization in the nearly flat and topological moiré minibands^{33,34}. However, the Hall resistance in the incipient Chern insulator state of the twisted bilayer graphene device of ref. ¹⁶ is not quantized even under a finite applied magnetic field, possibly owing to substantial twist-angle disorder³⁵.

Calculation for ABC-TLG/hBN moiré superlattice

A simple argument in ref. ⁶ shows that the valley Chern number must jump by 3 when switching the direction of the displacement field if the super-lattice gap does not close during this process. A direct numerical calculation gives $C=3$ for the $D<0$ side. Thus, owing to the interactions, a symmetry breaking state with spin-valley polarization is stabilized, and then a Chern insulator with $C=3$ is expected. We first improve our theoretical modelling of the band structure by adding various remote-hopping terms at the single-particle level and incorporating interaction effects, as shown in Extended Data Fig. 6. As we will show, $C=3$ is still robust in our more sophisticated model of single-particle band structure. Interaction effects turn out to be necessary to explain the reduction of the Chern number.

The ABC-TLG is modelled by a six-band model. We use the following parameters³⁶: $(v, \gamma_1, \gamma_2, \gamma_3, \gamma_4) = (2, 676, 380, 8.3, 260, 104)$ meV.

Then the Hamiltonian for the valley + is:

$$H_0(\mathbf{k}) = \begin{pmatrix} V/2 & v(k_x - ik_y) & -v_4(k_x - ik_y) & v_3(k_x + ik_y) & 0 & \gamma_2 \\ v(k_x + ik_y) & V/2 & 0 & 0 & 0 & 0 \\ -v_4(k_x + ik_y) & \gamma_1 & 0 & v(k_x - ik_y) & -v_4(k_x - ik_y) & v_3(k_x + ik_y) \\ v_3(k_x - ik_y) & 0 & v(k_x + ik_y) & 0 & \gamma_1 & 0 \\ 0 & 0 & -v_4(k_x + ik_y) & \gamma_1 & -\gamma/2 & v(k_x - ik_y) \\ \gamma_2 & 0 & v_3(k_x - ik_y) & 0 & v(k_x + ik_y) & -\gamma/2 \end{pmatrix}$$

The aligned hBN layer in the bottom provides a moiré hopping term which folds the original band structure to a small mini Brillouin zone. We use the same model of the moiré hopping term as in ref. ³⁷.

Using the above model we get the band structures shown in Fig. 4. For $\Phi_V = -25$ meV (where subscript V indicates ‘vertical’) we get a narrow valence band with Chern number $|C|=3$. We have tried to change the various hopping parameters and the potential difference. However, the Chern number is always equal to 3 and we conclude that $C=2$ cannot be reproduced at the single-particle level.

In Fig. 4b, we can see that although at each K-point the valence band is isolated from the band below, they overlap in energy (in other words, although there is a direct gap there is no indirect gap). More precisely, the system is in a compensated semimetal phase at the fully filled point. This is in agreement with the experimental measurement. On the $D < 0$ side, electrons are pushed away from the aligned bottom hBN layer. Therefore the moiré superlattice potential has a weaker effect and the superlattice gap is small.

Given that the valence band is not isolated from the remote band below, an interaction-induced self-energy can renormalize the band structure and maybe even the band topology. To incorporate this effect, we perform a self-consistent Hartree–Fock calculation by keeping only the valence band and the remote band. The interacting Hamiltonian is

$$H = \sum_{k,a,\sigma} \sum_m \xi_{a,m}(\mathbf{k}) c_{a,\sigma;m}^\dagger(\mathbf{k}) c_{a,\sigma;m}(\mathbf{k}) + \frac{1}{N} \sum_{\mathbf{q}} \sum_{\mathbf{k}_1, \mathbf{k}_2, a_1, \sigma_1, a_2, \sigma_2} c_{a_1, \sigma_1; m_1}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{a_2, \sigma_2; m_2}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a_2, \sigma_2; n_2}(\mathbf{k}_2) c_{a_1, \sigma_1; n_1}(\mathbf{k}_1) V(\mathbf{q}) \\ \lambda_{a_1; m_1 n_1}(\mathbf{k}_1, \mathbf{q}) \lambda_{a_2; m_2 n_2}(\mathbf{k}_2, -\mathbf{q})$$

where $a = \pm$ is the valley index and $m = 0, 1$ is the band index, labelling the valence band and the remote band. $\sigma = \uparrow, \downarrow$ is the spin index. $c_{a,\sigma;m}(\mathbf{k})$ is the creation operator corresponding to the band m for the valley a and spin σ . The two terms in the Hamiltonian are the kinetic term and the interaction term. $V(\mathbf{q})$ is the screened Coulomb interaction controlled by the renormalization factor of the dielectric constant. In the interaction we have included the form factors $\lambda_{a,mn}$ to incorporate the Berry curvature of the Bloch wavefunctions⁶. Then Hartree–Fock self-energy can be obtained from the self-consistent equations:

$$\Sigma_{a,mn}^H(\mathbf{k}) = \frac{1}{N} \sum_{a_2, \sigma; m'} \sum_{\mathbf{G}} \sum_{\mathbf{k}_2} c_{a_2, \sigma; m'}^\dagger(\mathbf{k}_2) V(\mathbf{G}) \lambda_{a,mn}(\mathbf{k}, \mathbf{G}) \lambda_{a_2; m' m}(\mathbf{k}_2, -\mathbf{G})$$

$$\Sigma_{a,mn}^F(\mathbf{k}) = -\frac{1}{N} \sum_{m'} \sum_{\mathbf{q}} c_{a\sigma; m'}^\dagger(\mathbf{k} + \mathbf{q}) c_{a\sigma; m'}(\mathbf{k} + \mathbf{q}) V(\mathbf{q}) \lambda_{a; m' n}(\mathbf{k}, \mathbf{q}) \\ \lambda_{a; mm'}(\mathbf{k} + \mathbf{q}, -\mathbf{q})$$

We solve these equations by iterating from zero initial values. Then we add the self-energies to calculate the new Chern number. The result is summarized in Fig. 4c. When the dielectric constant is large, the Chern number is 3, the same as the non-interacting case. For a fixed displacement field, increasing the interaction strength (decreasing the dielectric constant) can reduce the Chern number to 2 through a topological transition. For a large parameter region, $C=2$ is indeed expected, consistent with the current experiment.

Finally, we discuss the nature of the observed Chern insulator. At 1/4 filling, there is one particle per moiré unit cell. Within Hartree–Fock theory⁶, the most natural ground state is a spin- and valley-polarized Chern insulator with $C=2$, which is consistent with the transport measurement.

Because the two valleys are degenerate at zero magnetic field, domains can exist and cause the non-quantization of the Hall conductivity owing to the chiral edge mode in the domain boundary. Upon increasing the magnetic field, the valley Zeeman coupling can align the domains and lead to perfect quantization. In the current experiment,

the quantization is achieved at only 0.2 T, which suggests that the valley Zeeman coupling is large, consistent with previous theoretical calculations¹⁷.

The anomalous Hall effect in Fig. 3 strongly suggests that the valley is polarized. Although the simplest ansatz within mean field theory also requires spin polarization, the present transport measurement cannot rule out more exotic Chern insulator phases with the spins in a disordered (for example, ‘spin liquid’) or anti-ferromagnetic phase. Even for a simple spin-polarized scenario, non-trivial topological defects in spin space may have an important role. The skyrmion excitation carries charge $Q = 2e$ in the Chern insulator and may be the cheapest charge excitation (at small field the cheapest charge excitations may also be valley flips). In this case the activation gap is decided by the skyrmion gap. The existence of skyrmions may be reflected in a large g factor (skyrmions involve many spin flips) for response of activation gap to magnetic field. We leave it to future experiments to probe these possible interesting physics associated with the spin texture.

Insulating behaviour and extracted gap of the Chern insulator state

We measured the magnetic-field-dependent energy gap of the $\nu = 2$ state. Extended Data Fig. 5 shows the temperature dependent longitudinal resistivity ρ_{xx} and conductivity $\sigma_{xx} = \rho_{xx} / (\rho_{xx}^2 + \rho_{yx}^2)$, at different magnetic field B for the $\nu = 2$ state. For all magnetic fields, the temperature dependencies in Extended Data Fig. 5a, b exhibit typical behaviours of the quantum Hall insulator in that both ρ_{xx} and σ_{xx} decreases with decreasing temperature, which differs from a trivial insulator behaviour where the decrease in σ_{xx} is accompanied by a divergence in ρ_{xx} . This quantum-Hall-like behaviour persists at zero magnetic field. Extended Data Fig. 5c shows the energy gap at different magnetic fields, obtained by fitting the data to an Arrhenius activation model of $\sigma_{xx} \propto e^{-\Delta/2k_B T}$. The deviation from the Arrhenius behaviour at the low- and high-temperature limits is possibly due to the variable range hopping at low temperatures and strong thermal excitations at high temperatures. We extract a non-zero energy gap of about 2 K at $B = 0$. We can clearly see a continuous increase in gap size with increasing the magnetic field from $B = 0$, and the non-zero intercept at $B = 0$. It indicates that quantum Hall behaviour extends to zero magnetic field, consistent with the Chern insulator state.

Ferromagnetism and Chern insulator in a second device

Similar ferromagnetism and Chern insulator data have been observed in a second ABC-TLG/hBN device. Extended Data Fig. 7 shows the basic characterization of the second device (device II, Extended Data Fig. 7a). The moiré exists between the top hBN and ABC-TLG for device II (Extended Data Fig. 7b, opposite to that of device I in the main text). This leads to a non-trivial band at positive displacement fields and a trivial band at negative displacement fields. Extended Data Fig. 7c shows the Mott insulating states at 1/4 and 1/2 fillings of the trivial band for negative displacement fields, and weak resistance peaks of the non-trivial band for the positive displacement fields.

By tuning the gate voltages to the non-trivial side at the positive displacement field near 1/4 filling, we reproduced the main data of device I in the main text. Extended Data Fig. 8 shows the main results

for device II. At $D = 0.55 \text{ V nm}^{-1}$, an anomalous Hall hysteresis loop is clearly resolved. We show the n - and D -dependence of the anomalous Hall resistivity ρ_{xy}^{AH} and the coercive field B_c in Extended Data Fig. 8b. The fans of ρ_{xx} and ρ_{yx} in Extended Data Fig. 8c, d show the clear $\nu = 2$ state developing from 1/4 filling, as represented by the dashed lines. (The contact resistances are much larger in device II, which leads to much larger measurement noise. It also prevents us from measuring the magneto-transport at the lowest dilution fridge temperatures owing to the increased contact resistances at low temperature.)

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

32. Young, A. F. et al. Spin and valley quantum Hall ferromagnetism in graphene. *Nat. Phys.* **8**, 550–556 (2012).
33. Bultinck, N., Chatterjee, S. & Zaletel, M. P. Anomalous Hall ferromagnetism in twisted bilayer graphene. Preprint at <https://arxiv.org/abs/1901.08110> (2019).
34. Zhang, Y.-H., Mao, D. & Senthil, T. Twisted bilayer graphene aligned with hexagonal boron nitride: anomalous Hall effect and a lattice model. Preprint at <https://arxiv.org/abs/1901.08209> (2019).
35. Yoo, H. et al. Atomic and electronic reconstruction at the van der Waals interface in twisted bilayer graphene. *Nat. Mater.* **18**, 448–453 (2019).
36. Zhang, F., Sahu, B., Min, H. & MacDonald, A. H. Band structure of ABC-stacked graphene trilayers. *Phys. Rev. B* **82**, 035409 (2010).
37. Jung, J., Raoux, A., Qiao, Z. & MacDonald, A. H. Ab initio theory of moiré superlattice bands in layered two-dimensional materials. *Phys. Rev. B* **89**, 205414 (2014).

Acknowledgements We thank Y. Yu and M. Sui for measurement assistance, and acknowledge discussions with M. Zaletel, E. Altman, J. Jung and M. A. Kastner. G.C. and F.W. were supported as part of the Center for Novel Pathways to Quantum Coherence in Materials, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences. A.L.S. was supported by a National Science Foundation Graduate Research Fellowship and a Ford Foundation Predoctoral Fellowship. The work of E.J.F. and D.G.-G. on this project was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under contract number DE-AC02-76SF00515. Low-temperature infrastructure (dilution fridges) and cryostat support were funded in part by the Gordon and Betty Moore Foundation through grant number GBMF3429. Part of the sample fabrication was conducted at the Nano-fabrication Laboratory at Fudan University. Part of the measurement was performed in Oxford Instrument Nanoscience Shanghai Demo Laboratory. Y.Z. acknowledges financial support from National Key Research Program of China (grant numbers 2016YFA0300703, 2018YFA0305600), NSF of China (grant numbers U1732274, 11527805, 11425415 and 11421404), and the Strategic Priority Research Program of Chinese Academy of Sciences (grant number XDB30000000). Z.S. acknowledges support from National Key Research and Development Program of China (grant number 2016YFA0302001) and National Natural Science Foundation of China (grant number 11574204, 11774224), and additional support from a Shanghai talent programme. T.S. was supported by NSF grant DMR-1608505, and partially through a Simons Investigator Award from the Simons Foundation. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by the MEXT, Japan and the CREST (JPMJCR15F3), JST.

Author contributions F.W. and G.C. conceived the project. F.W., Y.Z., D.G.-G. and T.S. supervised the project. G.C. fabricated samples and performed transport characterizations at temperature above 1 K to first identify the $\nu = 2$ quantum Hall states. G.C., A.L.S. and E.J.F. performed ultralow temperature transport measurements. G.C., L.J., B.L., H.L. and Z.S. prepared TLG and performed near-field infrared and atomic force microscope measurements. K.W. and T.T. grew hBN single crystals. Y.-H.Z. and T.S. calculated the band structures and Chern numbers. G.C., A.L.S., E.J.F., Y.-H.Z., T.S., D.G.-G., Y.Z. and F.W. analysed the data. G.C., Y.-H.Z., T.S. and F.W. wrote the paper, with input from all authors.

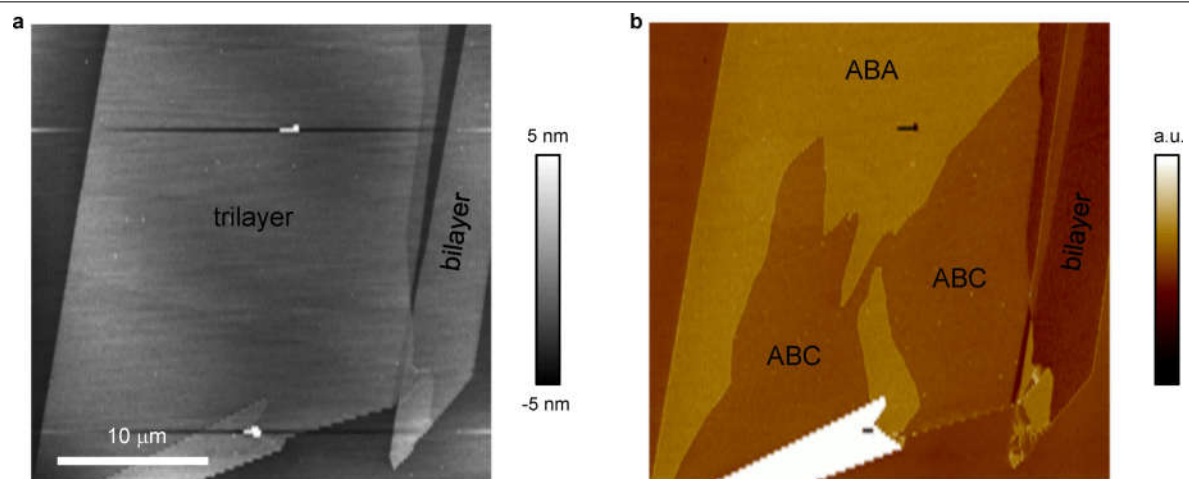
Competing interests The authors declare no competing interests.

Additional information

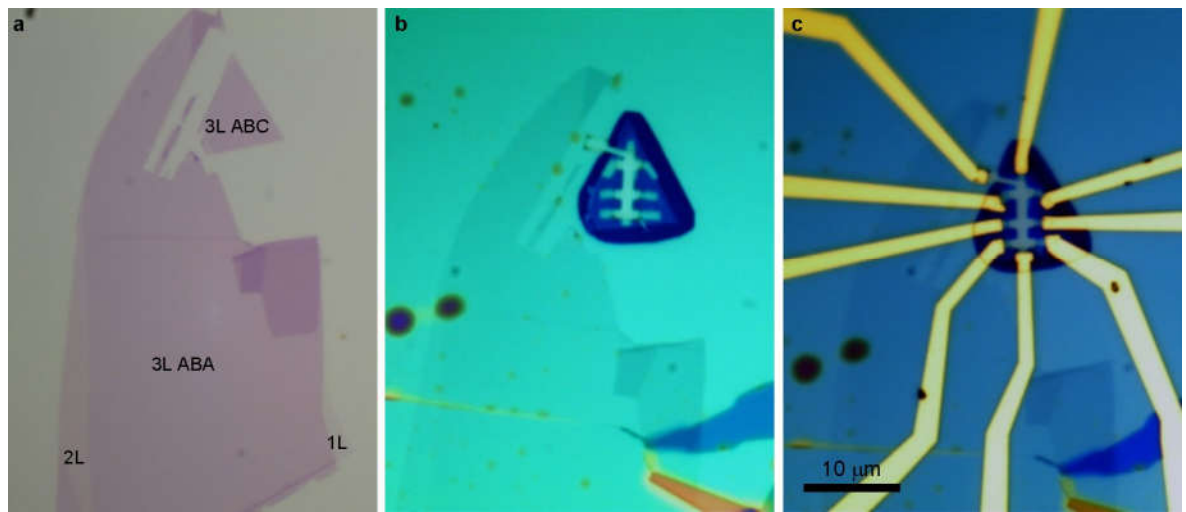
Correspondence and requests for materials should be addressed to D.G.-G., Y.Z. or F.W.

Peer review information Nature thanks Fan Zheng and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

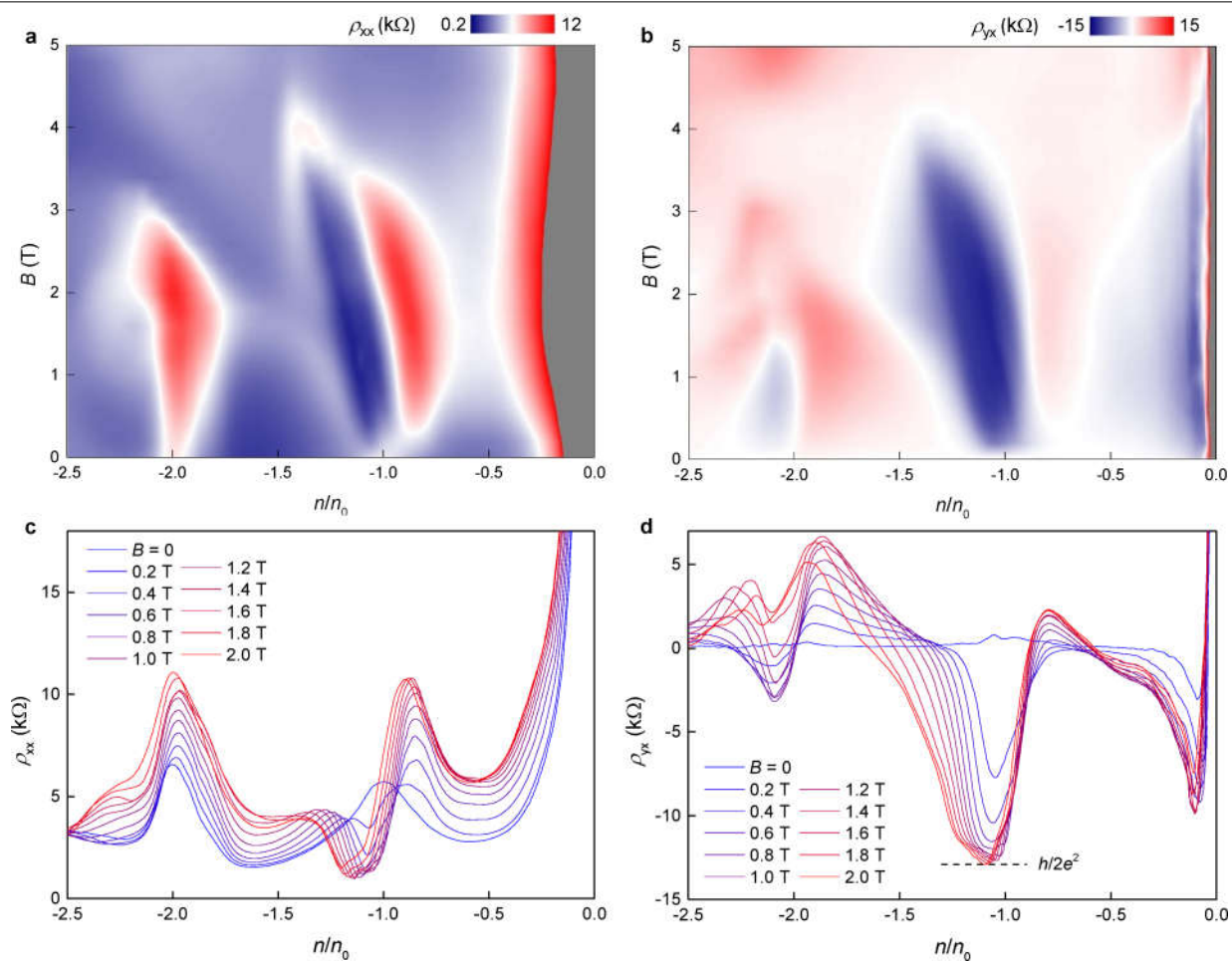
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Identification of ABC-TLG. **a**, Atomic force microscope topography image of an exfoliated TLG on SiO₂/Si. **b**, Near-field infrared image corresponding to **a**, showing that ABC-TLG has different contrast to ABA-TLG.

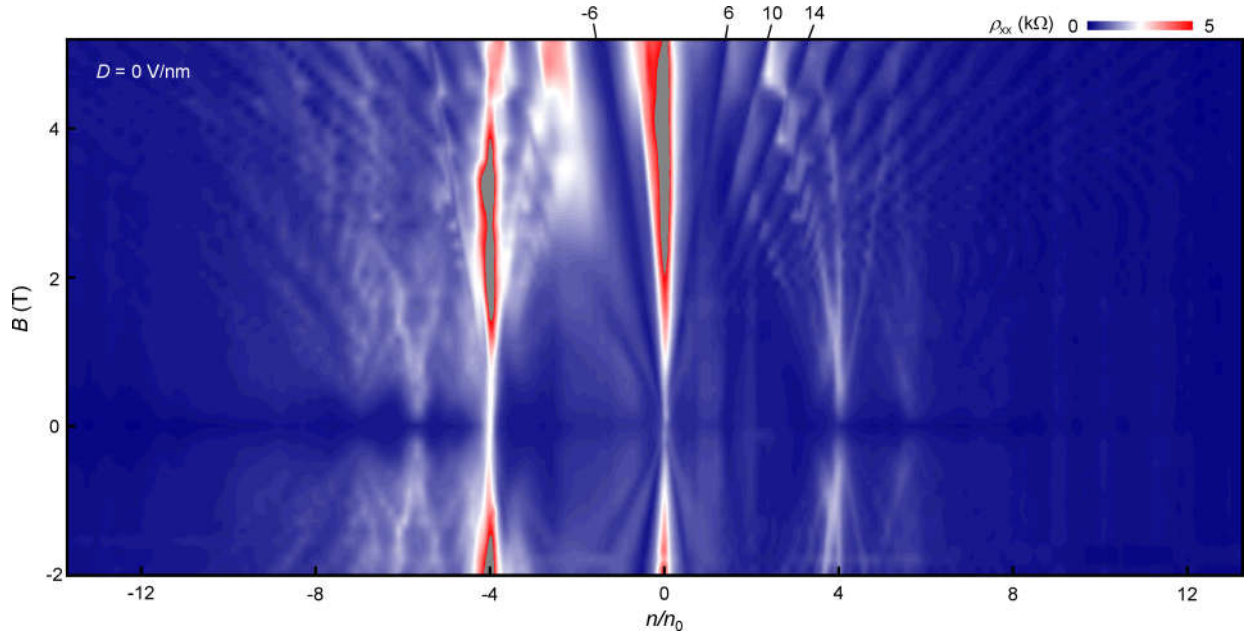


Extended Data Fig. 2 | Optical images of device I during fabrication. **a**, ABC-TLG is identified by near-field infrared spectroscopy and isolated by atomic force microscope tip. **b**, ABC-TLG is encapsulated by hBN and etched into Hall bar geometry. **c**, Final device with metal contacts and top and bottom gates.



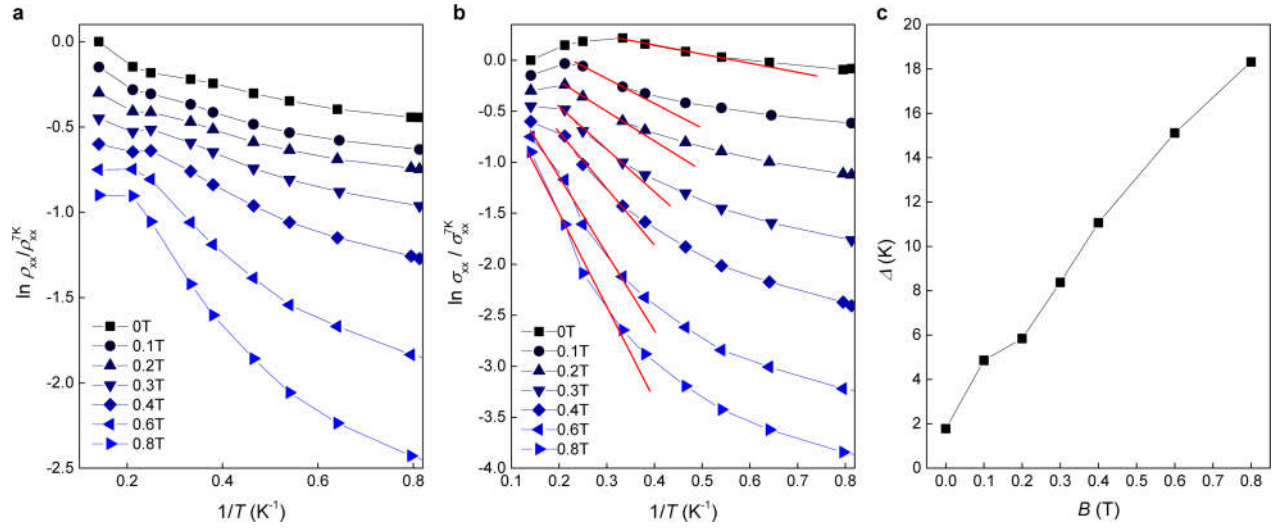
Extended Data Fig. 3 | Magneto-transport of the Chern insulator state at $T=1.5$ K. **a, b**, Colour plots of ρ_{xx} and ρ_{yx} as a function of carrier density and magnetic field at $D=-0.5$ V nm $^{-1}$ and $T=1.5$ K. The $\nu=2$ Chern insulator state is

well resolved at 1.5 K, which features a minimum for ρ_{xx} and a quantized ρ_{yx} emerges from 1/4 filling. **c, d**, Horizontal line cuts of **a** and **b**, respectively. ρ_{yx} shows quantized Hall resistance at finite magnetic field.



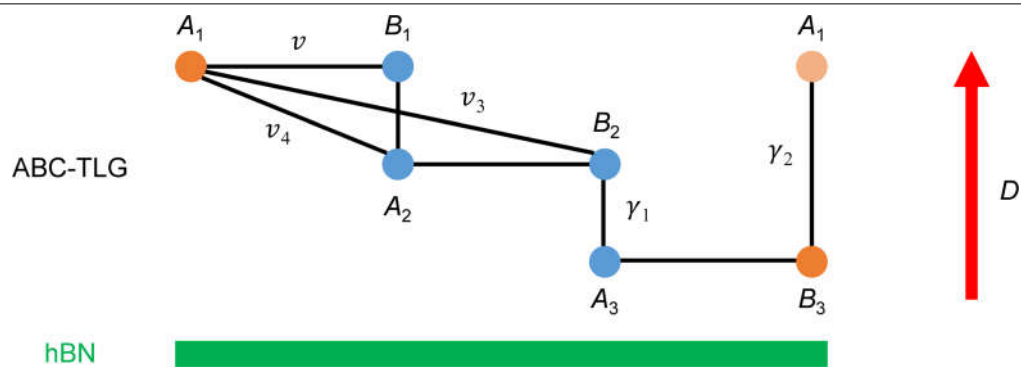
Extended Data Fig. 4 | Landau fan at $D=0$. Longitudinal resistivity ρ_{xx} (colour scale) as a function of carrier density and magnetic field at displacement field $D=0$. Clear Landau levels develop from the charge neutrality point and fully filled points at $D=0$, which is direct evidence of the high quality of the encapsulated ABC-TLG device described in the main text. The first resolved

quantum Hall state of the charge neutrality point is $\nu=6$. This Landau fan diagram establishes conclusively that we have ABC trilayer graphene in the hBN encapsulated device; it is completely different from the Landau fan diagram of ABA trilayer graphene (see ref.¹⁴).

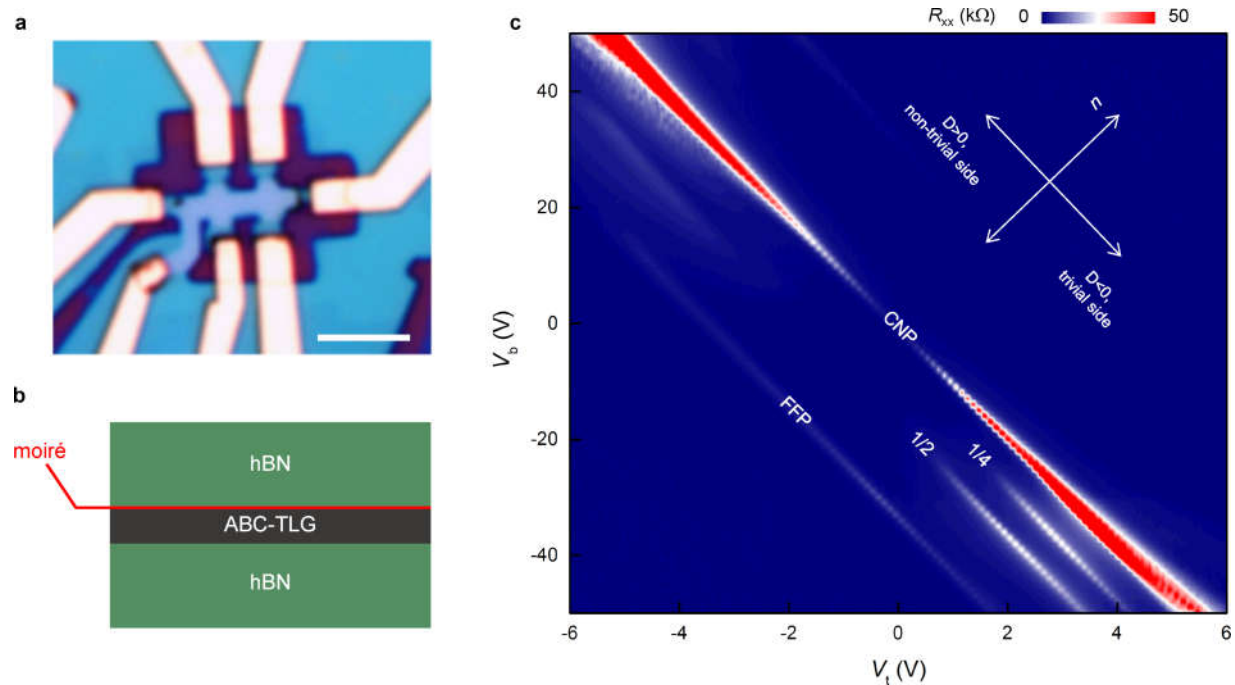


Extended Data Fig. 5 | Temperature dependence of the $\nu=2$ state. a–c, Arrhenius plot of longitudinal resistivity (a), conductivity (b) and the estimated gap at different magnetic field (c). A manual offset of -0.15 on the y axis is applied to each curve in a and b. The gap size in c is extracted from the linear fit of $\sigma_{xx} \propto e^{-\Delta/2k_B T}$ (red line) in b. We note that the Arrhenius plot is only valid for a limited temperature range, suggesting deviation from the thermal

activated behaviour at low temperatures. Therefore, the estimated gaps have relatively large uncertainty. However, the qualitative behaviour is robust: insulating behaviour is observed at all magnetic fields, and the quantized Hall insulator at finite magnetic field connects smoothly with the anomalous Hall insulator at zero magnetic field, supporting the identification of the state as a Chern insulator.

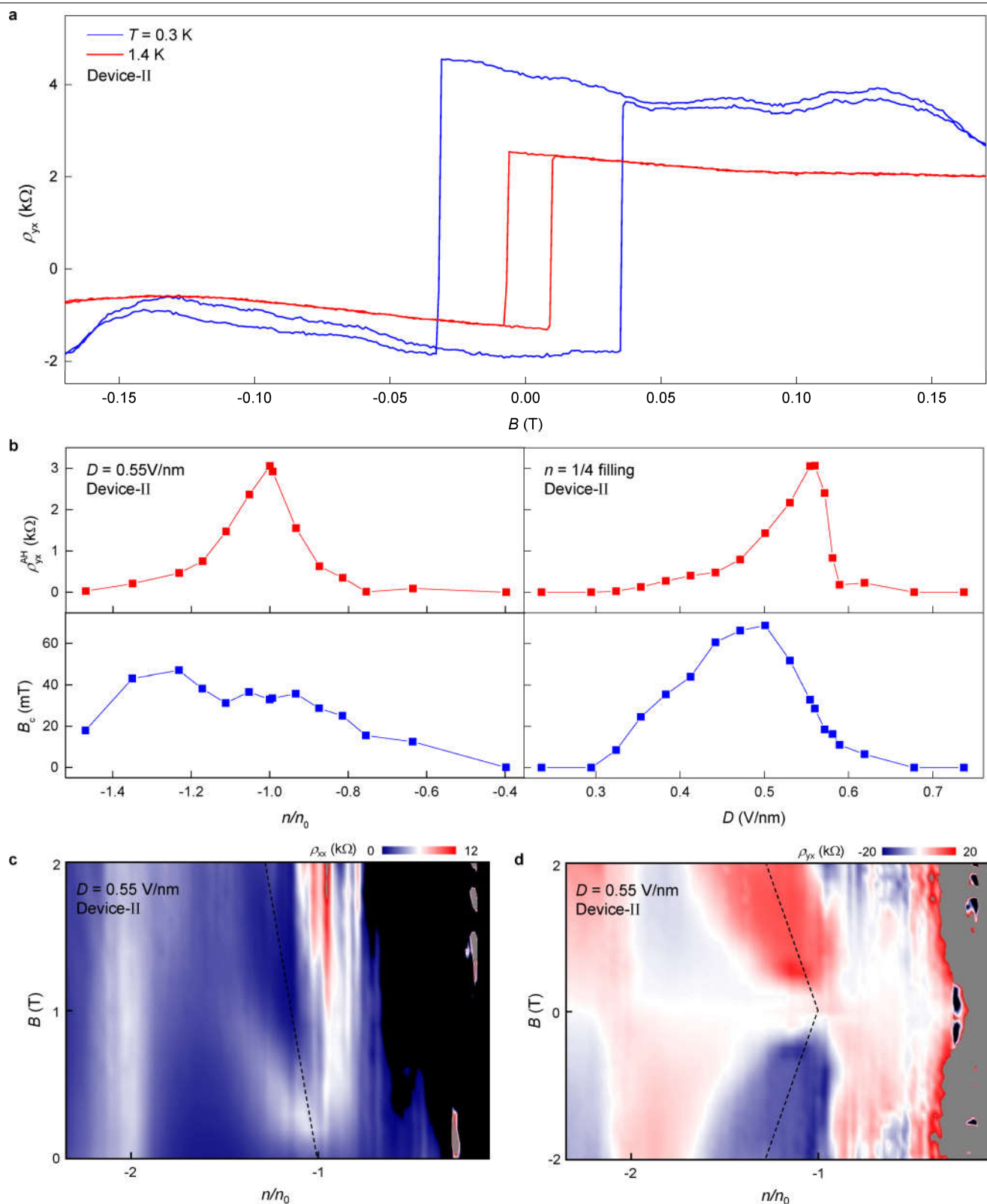


Extended Data Fig. 6 | Illustration of the ABC-TLG/hBN system. The bottom hBN layer is nearly aligned with the graphene layers whereas the one on top is not aligned. A and B refer to the two sublattices in each of the graphene layers.



Extended Data Fig. 7 | Basic characterizations of the second device (device II). **a**, Optical image of device II. The device is in a standard Hall bar geometry with top and bottom gates. The scale bar is $3\ \mu\text{m}$. **b**, Schematic of the moiré pattern existing between top hBN and ABC-TLG for device II. **c**, Two-

dimensional colour plot of R_{xx} as a function of V_t and V_b at $T = 5\ \text{K}$. The moiré exists between the top hBN and ABC-TLG for device II, opposite to that of device I in the main text. This leads to a non-trivial band at positive displacement fields and a trivial band at negative displacement fields.



Extended Data Fig. 8 | Reproducible Chern insulator data for device II. a, Ferromagnetic anomalous Hall effect at $1/4$ filling at 0.3 K and 1.4 K. **b,** The evolution of ρ_{yx}^{AH} and B_c as a function of doping (at $D = 0.55$ V nm $^{-1}$) and

displacement field (at $n = n_0$) at 1.1 K. **c, d,** Colour plot of ρ_{xx} and ρ_{yx} as a function of carrier density and magnetic field. Dashed lines represent the $\nu = 2$ state.

Ultrafast machine vision with 2D material neural network image sensors

<https://doi.org/10.1038/s41586-020-2038-x>

Received: 28 August 2019

Accepted: 17 January 2020

Published online: 4 March 2020

 Check for updates

Lukas Mennel^{1✉}, Joanna Symonowicz¹, Stefan Wachter¹, Dmitry K. Polyushkin¹, Aday J. Molina-Mendoza¹ & Thomas Mueller^{1✉}

Machine vision technology has taken huge leaps in recent years, and is now becoming an integral part of various intelligent systems, including autonomous vehicles and robotics. Usually, visual information is captured by a frame-based camera, converted into a digital format and processed afterwards using a machine-learning algorithm such as an artificial neural network (ANN)¹. The large amount of (mostly redundant) data passed through the entire signal chain, however, results in low frame rates and high power consumption. Various visual data preprocessing techniques have thus been developed^{2–7} to increase the efficiency of the subsequent signal processing in an ANN. Here we demonstrate that an image sensor can itself constitute an ANN that can simultaneously sense and process optical images without latency. Our device is based on a reconfigurable two-dimensional (2D) semiconductor^{8,9} photodiode^{10–12} array, and the synaptic weights of the network are stored in a continuously tunable photoresponsivity matrix. We demonstrate both supervised and unsupervised learning and train the sensor to classify and encode images that are optically projected onto the chip with a throughput of 20 million bins per second.

ANNs have achieved huge success as machine-learning algorithms in a wide variety of fields¹. The computational resources required to perform machine-learning tasks are very demanding. Accordingly, dedicated hardware solutions that provide better performance and energy efficiency than conventional computer architectures have become a major research focus. However, although much progress has been made in efficient neuromorphic processing of electrical^{13–16} or optical^{17–20} signals, the conversion of optical images into the electrical domain remains a bottleneck, particularly in time-critical applications. Imaging systems that mimic neuro-biological architectures may allow us to overcome these disadvantages. Much work has therefore been devoted to develop systems that emulate certain functions of the human eye²¹, including hemispherically shaped image sensors^{22,23} and preprocessing of visual data^{2–7}, for example, for image-contrast enhancement, noise reduction or event-driven data acquisition.

Here, we present a photodiode array that itself constitutes an ANN that simultaneously senses and processes images projected onto the chip. The sensor performs a real-time multiplication of the projected image with a photoresponsivity matrix. Training of the network requires setting the photoresponsivity value of each pixel individually. Conventional photodiodes that are based, for example, on silicon exhibit a fixed responsivity that is defined by the inner structure (chemical doping profile) of the device, and are thus not suitable for the proposed application. Other technologies such as photonic mixing²⁴ and metal-semiconductor-metal detectors²⁵ may, in principle, be suitable, but these device concepts bear additional challenges, such as nonlinear tunability of the photoresponse and bias-dependent (and hence weight-dependent) dark current. We have therefore chosen WSe₂—a 2D semiconductor—as the photoactive material. 2D semiconductors not only show strong light-matter interaction and excellent optoelectronic

properties^{8,9} but also offer the possibility of external tunability of the potential profile in a device—and hence its photosensitivity—by electrostatic doping using multi-gate electrodes^{10–12}. In addition, 2D materials technology has by now achieved a sufficiently high level of maturity to be employed in complex systems²⁶ and provides ease of integration with silicon readout/control electronics²⁷.

Figure 1a schematically illustrates the basic layout of the image sensor. It consists of N photoactive pixels arranged in a 2D array, with each pixel divided into M subpixels. Each subpixel is composed of a photodiode, which is operated under short-circuit conditions and under optical illumination delivers a photocurrent of $I_{mn} = R_{mn} E_n A = R_{mn} P_n$, where R_{mn} is the photoresponsivity of the subpixel, E_n and P_n denote the local irradiance and optical power at the n th pixel, respectively, and A is the detector area. $n = 1, 2, \dots, N$ and $m = 1, 2, \dots, M$ denote the pixel and subpixel indices, correspondingly. An integrated neural network and imaging array can now be formed by interconnecting the subpixels. Summing all photocurrents produced by the m th detector element of each pixel

$$I_m = \sum_{n=1}^N I_{mn} = \sum_{n=1}^N R_{mn} P_n \quad (1)$$

performs the matrix-vector product operation $\mathbf{I} = \mathbf{R}\mathbf{P}$, with $\mathbf{R} = (R_{mn})$ being the photoresponsivity matrix, $\mathbf{P} = (P_1, P_2, \dots, P_N)^T$ being a vector that represents the optical image projected onto the chip and $\mathbf{I} = (I_1, I_2, \dots, I_M)^T$ being the output vector. Provided that the R_{mn} value of each detector element can be set to a specific positive or negative value, various types of ANNs for image processing can be implemented (see Fig. 1c, d), with the synaptic weights being encoded in the photoresponsivity matrix. The expression ‘negative photoresponsivity’ is to be understood in this context as referring to the sign of the photocurrent.

¹Institute of Photonics, Vienna University of Technology, Vienna, Austria. ✉e-mail: lukas.mennel@tuwien.ac.at; thomas.mueller@tuwien.ac.at

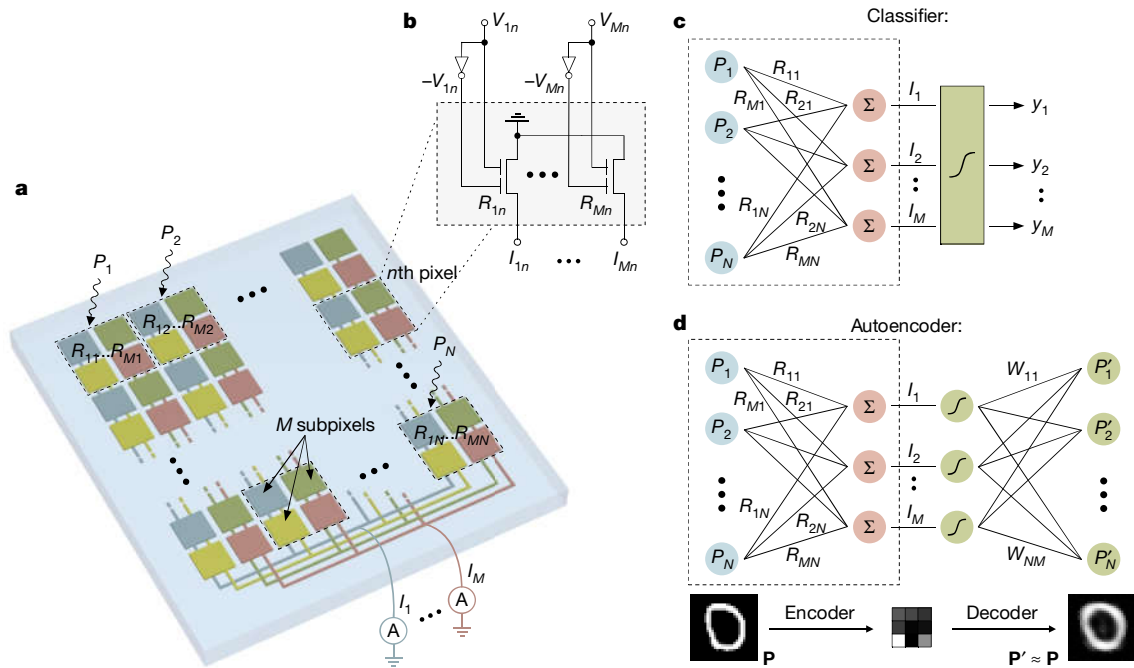


Fig. 1 | Imaging ANN photodiode array. **a**, Illustration of the ANN photodiode array. All subpixels with the same colour are connected in parallel to generate M output currents. **b**, Circuit diagram of a single pixel in the photodiode array. **c**, **d**, Schematics of the classifier (**c**) and the autoencoder (**d**). Below the

illustration of the autoencoder, shown is an example of encoding/decoding of a 28×28 pixel letter from the MNIST handwritten digit database. The original image is encoded to 9 code-layer neurons and then decoded back into an image.

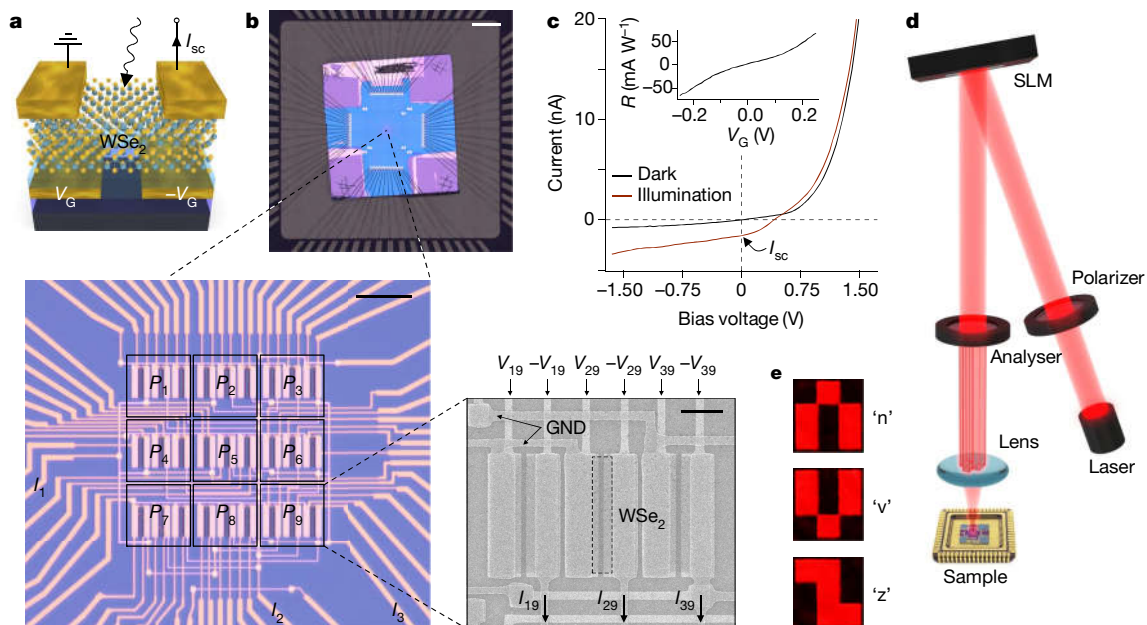


Fig. 2 | Implementation of the ANN photodiode array. **a**, Schematic of a single WSe_2 photodiode. The device is operated under short-circuit conditions and the photoresponsivity is set by supplying a voltage pair $V_G/-V_G$ to the bottom-gate electrodes. **b**, Macroscopic image of the bonded chip on the chip carrier. Scale bar, 2 mm. First magnification: microscope image of the photodiode array, which consists of 3×3 pixels. Scale bar, 15 μm . Second magnification: scanning electron microscopy image of one of the pixels. Each pixel consists of three WSe_2 photodiodes/subpixels with responsivities set by the gate voltages. Scale bar, 3 μm . GND, ground electrode. **c**, Current–voltage characteristic

curve of one of the photodetectors in the dark (blue line) and under optical illumination (red line). See also Extended Data Fig. 2a. The inset shows the gate-voltage tunability of the photoresponsivity. **d**, Schematic illustration of the optical setup. Laser light is linearly polarized by a wire-grid polarizer and reflected by a spatial light modulator (SLM). The reflected light is then filtered by an analyser (intensity modulation) and the resulting image is projected onto the photodiode array. **e**, Microscope images of the 3×3 pixel letters used for training/operation of the network.

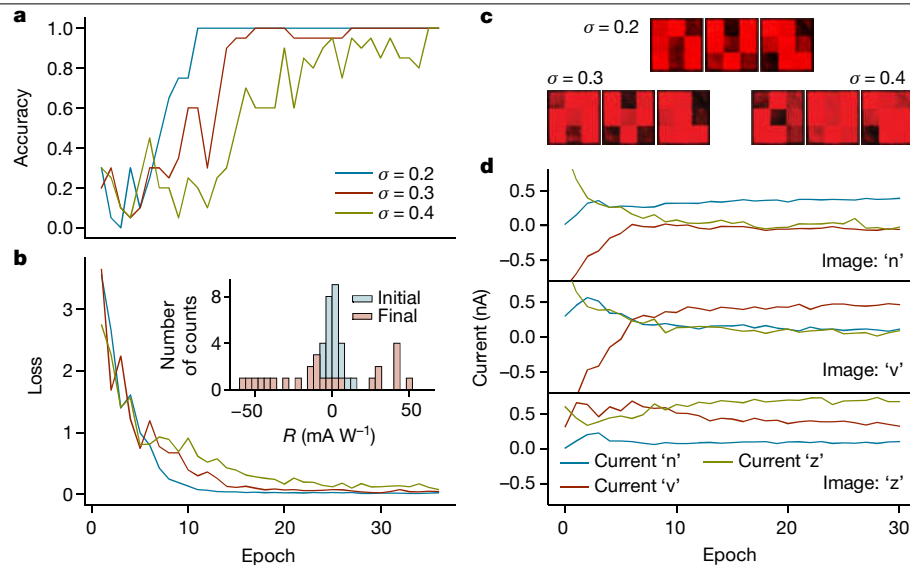


Fig. 3 | Device operation as a classifier. a, Accuracy of the classifier during training for varying artificial noise levels. An image is accurately predicted when the correct neuron has the largest activation. **b**, Loss function for different noise levels during training. The inset shows the initial and final responsivity distributions for $\sigma = 0.2$. **c**, Microscope images of the projected letters with different random noise levels. The complete dataset obtained over 30 epochs of training is shown in Extended Data Fig. 7. **d**, Average currents for

each epoch for each projected letter, measured during training with a noise level of $\sigma = 0.2$. Each graph shows the results of a separate experiment, in which the letters 'n' (top), 'v' (middle) and 'z' (bottom) are projected onto the chip, and three currents—corresponding to 'n' (blue), 'v' (red) and 'z' (green)—are measured. In the top graph, for example, the output that corresponds to the letter 'n' (current 'n') is the highest, so the ANN determines that the projected letter is 'n'.

We implemented two types of ANNs: a classifier and an autoencoder. Figure 1c shows a schematic of the classifier. Here, the array is operated as a single-layer perceptron, together with nonlinear activation functions that are implemented off-chip. This type of ANN represents a supervised learning algorithm that is capable of classifying images \mathbf{P} into different categories \mathbf{y} . An autoencoder (Fig. 1d) is an ANN that can learn, in an unsupervised training process, an efficient representation (encoding) for a set of images \mathbf{P} . Along with the encoder, a decoder is trained to attempt to reproduce at its output the original image, $\mathbf{P}' \approx \mathbf{P}$, from the compressed data. Here the encoder is formed by the photodiode array itself and the decoder by external electronics.

Having presented the operational concept of our network, we now come to an actual device implementation. We used a few-layer WSe_2 crystal with a thickness of about 4 nm to form lateral p–n junction photodiodes, using split-gate electrodes (with a ~300-nm-wide gap) that couple to two different regions of the 2D semiconductor channel (Fig. 2a)^{10–12}. WSe_2 was chosen because of its ambipolar conduction behaviour and excellent optoelectronic properties. Biasing one gate electrode at V_G and the other at $-V_G$ enables adjustable (trainable) responsivities between -60 and $+60 \text{ mA W}^{-1}$, as shown in Fig. 2c. This technology was then used to fabricate the photodiode array shown in Fig. 2b, which consists of 27 detectors with good uniformity, tunability and linearity (see Extended Data Figs. 1, 2b). The devices were arranged to form a 3×3 imaging array ($N = 9$) with a pixel size of about $17 \times 17 \mu\text{m}^2$ and with three detectors per pixel ($M = 3$). The short-circuit photocurrents I_{sc} produced by the individual devices under optical illumination were summed according to Kirchhoff's law by hard-wiring the devices in parallel, as depicted in Fig. 1b. The sample fabrication is explained in Methods, and a schematic of the entire circuit is provided in Extended Data Fig. 3. Each device was supplied with a pair of gate voltages, V_G and $-V_G$, to set its responsivity individually. For training and testing of the chip, optical images were projected using the setup shown in Fig. 2d (for details, see Methods). Unless otherwise stated, all measurements were performed using light with a wavelength of 650 nm and with a maximum irradiance of about 0.1 W cm^{-2} . Despite its small size, such a network is sufficient for the proof-of-principle demonstration

of several machine-learning algorithms. In particular, we performed classification, encoding, and denoising of the stylized letters 'n', 'v' and 'z' depicted in Fig. 2e. Scaling the network to larger dimensions is conceptually straightforward and remains a mainly technological task.

To test the functionality of the photodiode array, we first operated it as a classifier (Fig. 1c) to recognize the letters 'n', 'v' and 'z'. During each training epoch we optically projected a set of $S = 20$ randomly chosen letters. Gaussian noise (with standard deviation of $\sigma = 0.2, 0.3$ and 0.4 ; Fig. 3c) was added to augment the input data²⁸. In this supervised learning example, we chose one-hot encoding, in which each of the three letters activates a single output node/neuron. As activation function (the nonlinear functional mapping between the inputs and the output of a node) for the M photocurrents we chose the softmax function $\phi_m(I) = e^{I_m \xi} / \sum_{k=1}^M e^{I_k \xi}$ (a common choice for one-hot encoding), where $\xi = 10^{10} \text{ A}^{-1}$ is a scaling factor that ensures that the full value range of the activation function is accessible during training. As a loss/cost function (the function to be minimized during training) we used the cross-entropy $\mathcal{L} = -\frac{1}{M} \sum_{m=1}^M y_m \log [\phi_m(I)]$, where y_m is the label and $M = 3$ is the number of classes. The activations of the output neurons represent the probabilities for each of the letters. The initial values of the responsivities were randomly chosen from a Gaussian distribution, as suggested in ref.²⁹, and were different for the supervised- and unsupervised-learning demonstrations. The responsivities were updated after every epoch by backpropagation³⁰ of the gradient of the loss function

$$R_{mn} \rightarrow R_{mn} - \frac{\eta}{S} \sum_{\mathbf{P}} \nabla_{R_{mn}} \mathcal{L} \quad (2)$$

with learning rate $\eta = 0.1$. A detailed flow chart of the training algorithm is presented in Extended Data Fig. 4d.

In Fig. 3a, b the accuracy and loss are plotted over 35 training epochs. The loss is decreasing quickly for all noise levels and reaches a minimum after 15, 20 and 35 epochs for $\sigma = 0.2$, $\sigma = 0.3$ and $\sigma = 0.4$, respectively. The accuracy reaches 100% for all noise levels, with faster convergence for less noise. In Fig. 3d we show the mean currents for each of the three

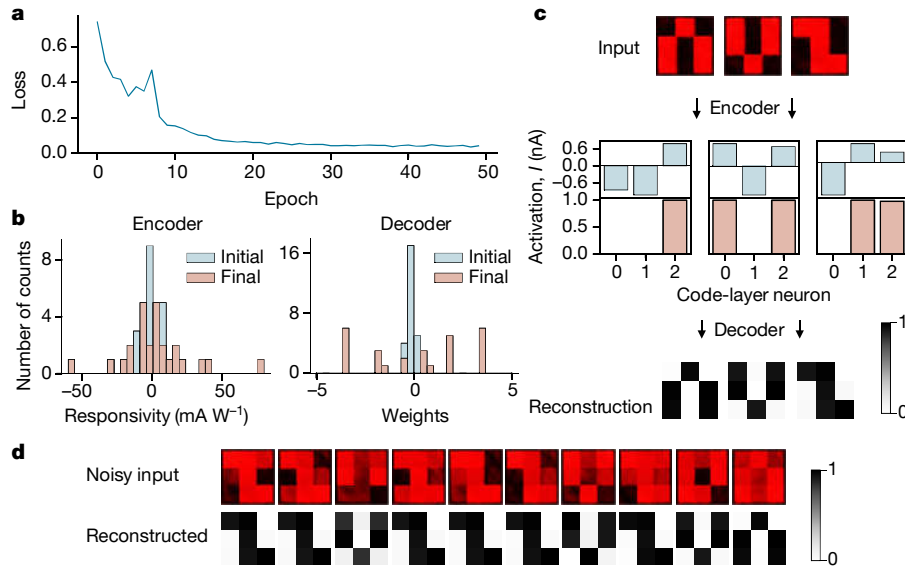


Fig. 4 | Device operation as an autoencoder. a, Loss of the autoencoder during training. The complete dataset of 30 epochs of training is given in Extended Data Fig. 7. **b**, Responsivity and weight distributions before (initial) and after (final) training. **c**, Autoencoding of noise-free letters. The encoder translates

the projected images into a current code, which is converted by the nonlinearity into a binary activation code and finally reconstructed into an image by the decoder. **d**, Randomly chosen noisy inputs ($\sigma=0.3$) and the corresponding reconstructions after autoencoding.

letters during each epoch for $\sigma=0.2$ (see Extended Data Fig. 5c, d for the other cases). The currents become well separated after about 10 epochs, with the highest current corresponding to the label of the projected letter. The inset in Fig. 3b shows histograms for the (randomly chosen) initial and final responsivity values for $\sigma=0.2$ (see also Extended Data Fig. 5a, b). The robustness and reliability of the classification results of the analogue vision sensor were verified by comparison of the accuracy and loss with computer simulations of a digital system with the same architecture and learning scheme (Extended Data Fig. 6).

Next, we demonstrate encoding of image patterns with our device operating as an autoencoder (Fig. 1d). We chose logistic (sigmoid) activation functions for the code neurons $\phi_m(I_m) = (1 + e^{-I_m/\xi})^{-1}$, again with $\xi = 10^{10} \text{ A}^{-1}$ as a scaling factor, as well as for the output neurons $P'_n = \phi_n(z_n) = (1 + e^{-z_n})^{-1}$, where $z_n = \sum_{m=1}^N W_{nm} \phi_m(I_m)$ and W_{nm} denotes the weight matrix of the decoder. We used the mean-square loss function $\mathcal{L} = \frac{1}{2} \|\mathbf{P} - \mathbf{P}'\|^2$, which depends on the difference between the original and reconstructed images. The responsivities were again trained by backpropagation of the loss according to equation (2), with a noise level of $\sigma = 0.15$. Along with the encoder responsivities, the weights of the decoder W_{nm} were trained. As shown in Fig. 4a, the loss steeply decreases within the first 10 training epochs and then slowly converges to a final value after about 30 epochs. The initial and final responsivities/weights of the encoder/decoder are shown in Fig. 4b and Extended Data Fig. 8, and the coded representations for each letter are depicted in Fig. 4c. Each projected letter delivers a unique signal pattern at the output. A projected 'n' delivers negative currents to code-layer neurons 1 and 2 and a positive current to code-layer neuron 3. After the sigmoid function, this causes only code-layer neuron 3 to deliver a sizeable signal. The letters 'v' and 'z' activate two code-layer neurons: 'v', code-layer neurons 0 and 2; 'z', code-layer neurons 1 and 2. The decoder transforms the coded signal back into an output that correctly represents the input. To test the fault tolerance of the autoencoder, we projected twice as noisy ($\sigma = 0.3$) images. Not only did the autoencoder interpret the inputs correctly, but the reconstructions were considerably less noisy (Fig. 4d).

As image sensing and processing are both performed in the analogue domain, the operation speed of the system is limited only by physical processes involved in the photocurrent generation³¹. As a result, image

recognition and encoding occur in real time with a rate that is orders of magnitude higher than what can be achieved conventionally. To demonstrate the high-speed capabilities of the sensor, we performed measurements with a 40-ns pulsed laser source (522 nm, $\sim 10 \text{ W cm}^{-2}$). The photodiode array was operated as a classifier and trained beforehand, as discussed above. We subsequently projected two letters ('v' and 'n') and measured the time-resolved currents of the two corresponding channels. In Fig. 5 we plot the electric output pulses, which demonstrate correct pattern classification within $\sim 50 \text{ ns}$. The system is thus capable of processing images with a throughput of 20 million bins per second. This value is limited only by the 20-MHz bandwidth of the used amplifiers, and substantially higher rates are possible. Such a network may hence provide new opportunities for ultrafast machine vision. It may also be employed in ultrafast spectroscopy for the detection and classification of spectral events. We also note that the operation of the vision sensor is self-powered (photovoltaic device) and electrical energy is consumed only during training.

Let us now comment on the prospects for scalability. In our present implementation the weights of the ANN are stored in an external memory and supplied to each detector via cabling. Scaling will require storing the weights locally. This could be achieved, for example, by using ferroelectric gate dielectrics or by employing floating gate devices^{32–34}. To demonstrate the feasibility of the latter approach, we present in Extended Data Fig. 9 a floating split-gate photodetector. Once set, this detector 'remembers' its responsivity value and delivers a photocurrent of adjustable sign/magnitude. During training, each detector could then be addressed by its column and row, using the standard infrastructure of active pixel cameras.

Another important question is the number of required subpixels M . As shown in the example in Fig. 1d, a segmentation of each pixel into 3×3 subpixels may be adequate for some applications. Given the exponential increase of network complexity with M , increasing the segmentation to 6×6 subpixels would already result in a very powerful ANN with a manageable number of 36 analogue outputs. We propose that such a network may also be trained as a binary-hashing³⁵ autoencoder, eliminating the need for analogue-to-digital conversion. Binary hashing encodes each feature into a binary code of the output signal, which means that a 36-bit digital output allows as many as $2^{36} - 1 \approx 7 \times 10^{10}$

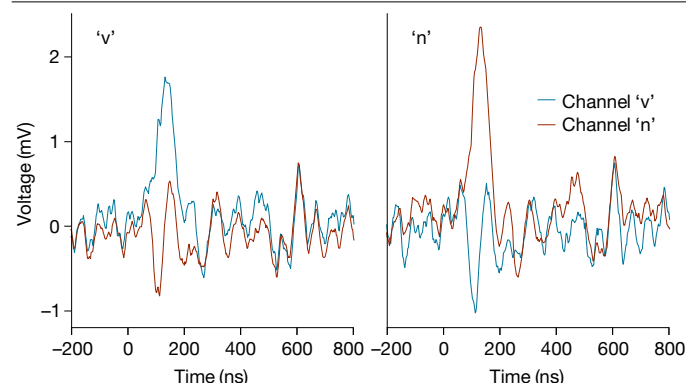


Fig. 5 | Ultrafast image recognition. Projection of two different letters, 'v' and 'n', with a duration of 40 ns, leads to distinct output voltages of the labelled channels.

encodable features. The implementation of an analogue deep-learning network becomes feasible by converting the photocurrents into voltages that are then fed into a memristor crossbar. We finally remark that besides on-chip training, demonstrated here, the network can also be trained off-line using computer simulations, and the predetermined photoresponsivity matrix is then transferred to the device.

In conclusion, we have presented an ANN vision sensor for ultrafast recognition and encoding of optical images. The device concept is easily scalable and provides various training possibilities for ultrafast machine vision applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2038-x>.

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Mead, C. A. & Mahowald, M. A. A silicon model of early visual processing. *Neural Netw.* **1**, 91–97 (1988).
3. Lichtsteiner, P., Posch, C. & Delbruck, T. A. 128×128 120 dB 15 μ s latency asynchronous temporal contrast vision sensor. *IEEE J. Solid-State Circuits* **43**, 566–576 (2008).
4. Cottini, N., Gottardi, M., Massari, N., Passerone, R. & Smilansky, Z. A. 33 μ W 64×64 pixel vision sensor embedding robust dynamic background subtraction for EVENT detection and scene interpretation. *IEEE J. Solid-State Circuits* **48**, 850–863 (2013).
5. Kyuma, K. et al. Artificial retinas—fast, versatile image processors. *Nature* **372**, 197–198 (1994).
6. Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B. & Delbruck, T. Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. *Proc. IEEE* **102**, 1470–1484 (2014).
7. Zhou, F. et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* **14**, 776–782 (2019).

8. Manzeli, S., Ovchinnikov, D., Pasquier, D., Yazyev, O. V. & Kis, A. 2D transition metal dichalcogenides. *Nat. Rev. Mater.* **2**, 17033 (2017).
9. Mueller, T. & Malic, E. Exciton physics and device application of two-dimensional transition metal dichalcogenide semiconductors. *npj 2D Mater. Appl.* **2**, 29 (2018).
10. Pospischil, A., Furchi, M. M. & Mueller, T. Solar-energy conversion and light emission in an atomic monolayer p–n diode. *Nat. Nanotechnol.* **9**, 257–261 (2014).
11. Baugher, B. W., Churchill, H. O., Yang, Y. & Jarillo-Herrero, P. Optoelectronic devices based on electrically tunable p–n diodes in a monolayer dichalcogenide. *Nat. Nanotechnol.* **9**, 262–267 (2014).
12. Ross, J. S. et al. Electrically tunable excitonic light-emitting diodes based on monolayer WSe₂ p–n junctions. *Nat. Nanotechnol.* **9**, 268–272 (2014).
13. Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).
14. Sheridan, P. M. et al. Sparse coding with memristor networks. *Nat. Nanotechnol.* **12**, 784–789 (2017).
15. Li, C. et al. Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **1**, 52–59 (2018).
16. Kim, K. H. et al. A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications. *Nano Lett.* **12**, 389–395 (2012).
17. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
18. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
19. Hamerly, R., Bernstein, L., Sluuds, A., Soljačić, M. & Englund, D. Large-scale optical neural networks based on photoelectric multiplication. *Phys. Rev. X* **9**, 021032 (2019).
20. Psaltis, D., Brady, D., Gu, X. G. & Lin, S. Holography in artificial neural networks. *Nature* **343**, 325–330 (1990).
21. Kolb, H. How the retina works: much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. *Am. Sci.* **91**, 28–35 (2003).
22. Jeong, K.-H., Kim, J. & Lee, L. P. Biologically inspired artificial compound eyes. *Science* **312**, 557–561 (2006).
23. Choi, C. et al. Human eye-inspired soft optoelectronic device using high-density MoS₂-graphene curved image sensor array. *Nat. Commun.* **8**, 1664 (2017).
24. Schwarte, R. et al. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). In *Proc. SPIE Sensors, Sensor Systems, and Sensor Data Processing* Vol. 3100, 245–253 (SPIE, 1997).
25. Sugeta, T., Urisu, T., Sakata, S. & Mizushima, Y. Metal-semiconductor-metal photodetector for high-speed optoelectronic circuits. *Jpn. J. Appl. Phys.* **19**, 459 (1980).
26. Wachter, S., Polyushkin, D. K., Bethge, O. & Mueller, T. A microprocessor based on a two-dimensional semiconductor. *Nat. Commun.* **8**, 14948 (2017).
27. Goossens, S. et al. Broadband image sensor array based on graphene–CMOS integration. *Nat. Photon.* **11**, 366–371 (2017).
28. Bishop, C. M. Training with noise is equivalent to Tikhonov regularization. *Neural Comput.* **7**, 108–116 (1995).
29. Bengio, Y. in *Neural Networks: Tricks of the Trade* Vol. 7700 (eds Montavon G. et al.) 437–478 (Springer, 2012).
30. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
31. Massicotte, M. et al. Dissociation of two-dimensional excitons in monolayer WSe₂. *Nat. Commun.* **9**, 1633 (2018).
32. Li, D. et al. Two-dimensional non-volatile programmable p–n junctions. *Nat. Nanotechnol.* **12**, 901–906 (2017).
33. Lv, L. et al. Reconfigurable two-dimensional optoelectronic devices enabled by local ferroelectric polarization. *Nat. Commun.* **10**, 3331 (2019).
34. Bertolazzi, S., Krasnozhan, D. & Kis, A. Nonvolatile memory cells based on MoS₂/graphene heterostructures. *ACS Nano* **7**, 3246–3252 (2013).
35. Salakhutdinov, R. & Hinton, G. Semantic hashing. *Int. J. Approx. Reason.* **50**, 969–978 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Device fabrication

The fabrication of the chip followed the procedure described in ref.²⁶. As a substrate we used a silicon wafer, coated with 280-nm-thick SiO₂. First, we prepared a bottom metal layer by writing a design with electron-beam lithography (EBL) and evaporating Ti/Au (3 nm/30 nm). Secondly, we deposited a 30-nm-thick Al₂O₃ gate oxide using atomic layer deposition. Via holes through the Al₂O₃ isolator, which were necessary for the connections between the top and bottom metal layers, were defined by EBL and etched with a 30% solution of KOH in deionized water. Thirdly, we mechanically exfoliated a $\sim 70 \times 120 \mu\text{m}^2$ WSe₂ flake from a bulk crystal (from HQ Graphene) and transferred it onto the desired position on the sample by an all-dry viscoelastic stamping method³⁶. The crystal thickness (about six monolayers, or ~ 4 nm) was estimated from the contrast under which it appears in an optical microscope. Next, we separated 27 pixels from the previously transferred WSe₂ sheet by defining a mask with EBL and reactive ion etching with Ar/SF₆ plasma. Mild treatment with reactive ion etching oxygen plasma allowed the removal of the crust from the surface of the polymer mask that appeared during the preceding etching step. Then, a top metal layer was added by another EBL process and Ti/Au (3 nm/32 nm) evaporation. We confirmed the continuity and solidity of the electrode structure by scanning electron microscopy and electrical measurements. Finally, the sample was mounted in a 68-pin chip carrier and wire-bonded.

Experimental setup

Schematics of the experimental setup are shown in Fig. 2d and Extended Data Fig. 4a–c. Light from a semiconductor laser (650 nm wavelength) was linearly polarized before it illuminated a spatial light modulator (SLM; Hamamatsu), operated in intensity-modulation mode. On the SLM, the letters were displayed and the polarization of the light was rotated depending on the pixel value. A linear polarizer with its optical axis oriented normal to the polarization direction of the incident laser light functioned as an analyser. The generated optical image was then projected onto the sample using a 20 \times microscope objective with long working distance (Mitutoyo). Pairs of gate voltages were supplied to each of the detectors individually using a total of 54 digital-to-analogue

converters (National Instruments, NI-9264) and the three output currents were measured by source meters (Keithley, 2614B). For time-resolved measurements, a pulsed laser source emitting ~ 40 -ns-long pulses at 522 nm wavelength was used. The output current signals were amplified with high-bandwidth (20 MHz) transimpedance amplifiers (Femto) and the output voltages were recorded with an oscilloscope (Keysight). For the time-resolved measurements, the analyser was removed and the SLM was operated in phase-only mode to achieve higher illumination intensities ($\sim 10 \text{ W cm}^{-2}$). The phase-only Fourier transforms of the projected images were calculated using the Gerchberg–Saxton algorithm³⁷. For reliable and hysteresis-free operation, the vision sensor was placed in a vacuum chamber ($\sim 10^{-6}$ mbar). Alternatively, a protective dielectric encapsulation layer may be employed to isolate the two-dimensional semiconductor from the environment.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

36. Castellanos-Gomez, A. et al. Deterministic transfer of two-dimensional materials by all-dry viscoelastic stamping. *2D Mater.* **1**, 011002 (2014).

37. Gerchberg, R. W. & Saxton, W. O. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik* **35**, 237–246 (1972).

Acknowledgements We thank B. Limbacher for discussions and A. Kleinl, M. Paur and F. Dona for technical assistance. We acknowledge financial support by the Austrian Science Fund FWF (START Y 539-N16) and the European Union (grant agreement number 785219 Graphene Flagship and Marie Skłodowska-Curie Individual Fellowship OPTOVanderWAALS, grant ID 791536).

Author contributions T.M. conceived the experiment. L.M. designed and built the experimental setup, programmed the machine-learning algorithms, carried out the measurements and analysed the data. J.S. fabricated the ANN vision sensor. S.W. and D.K.P. contributed to the sample fabrication. A.J.M.-M. fabricated and characterized the floating-gate detector. L.M., J.S. and T.M. prepared the manuscript. All authors discussed the results and commented on the manuscript.

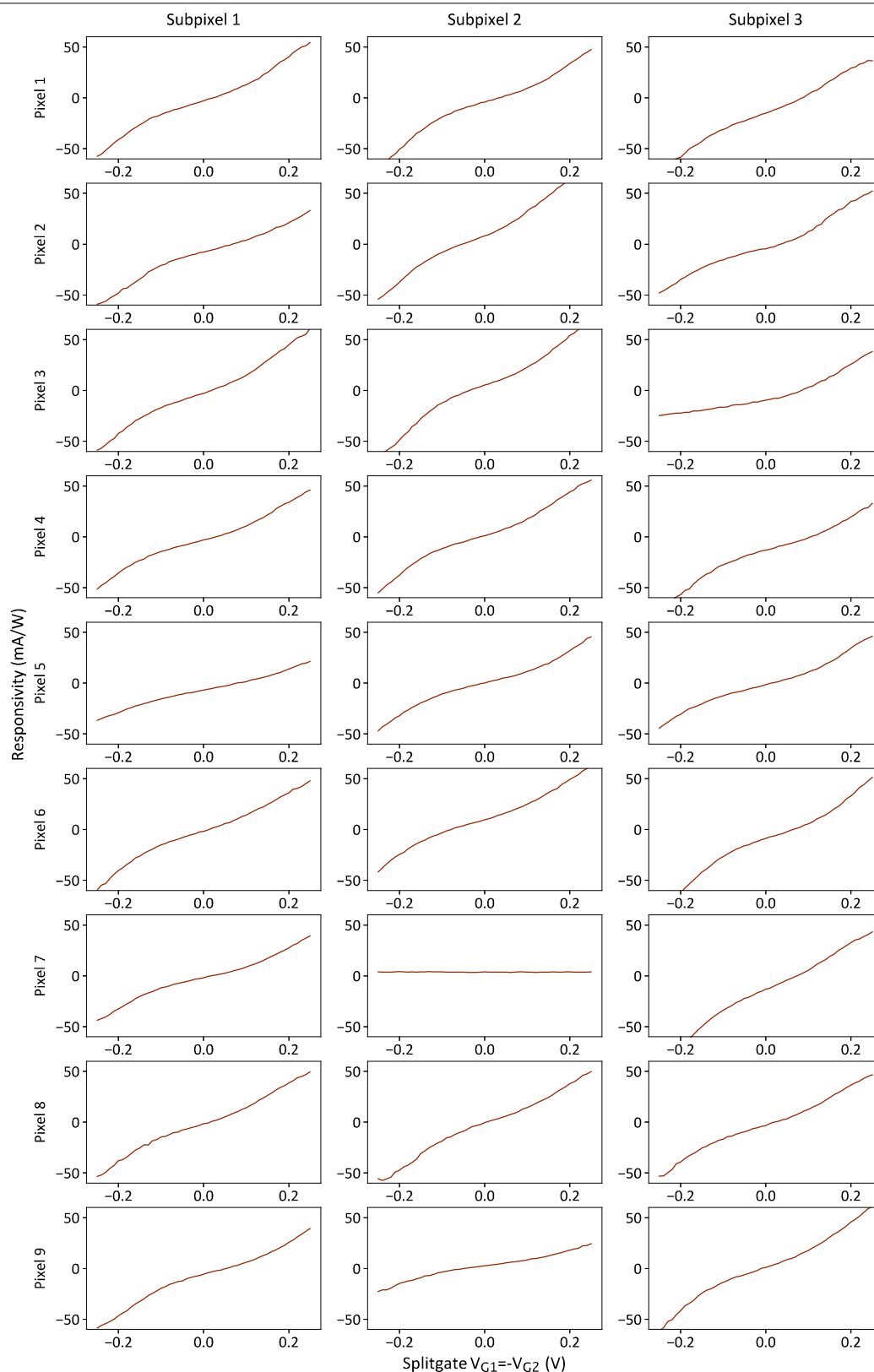
Competing interests The authors declare no competing interests.

Additional information

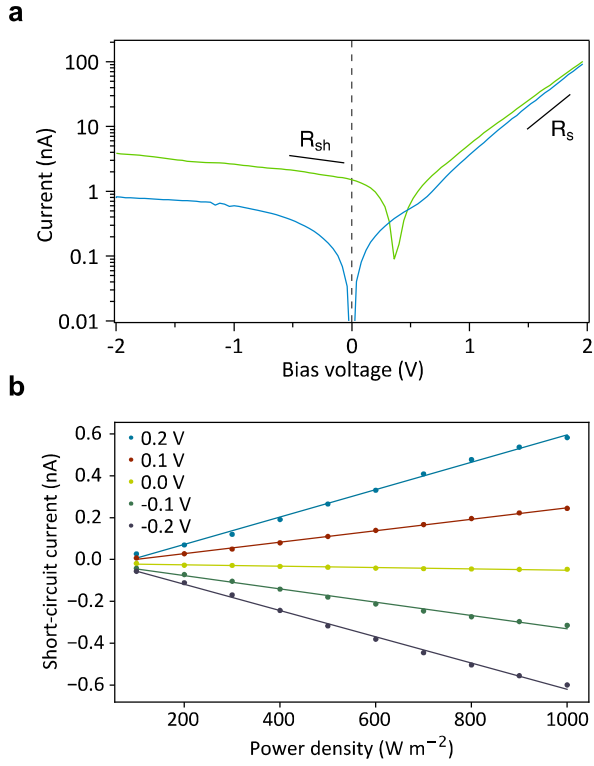
Correspondence and requests for materials should be addressed to L.M. or T.M.

Peer review information *Nature* thanks Yang Chai, Frank Koppens and Sangyoun Lee for their contribution to the peer review of this work.

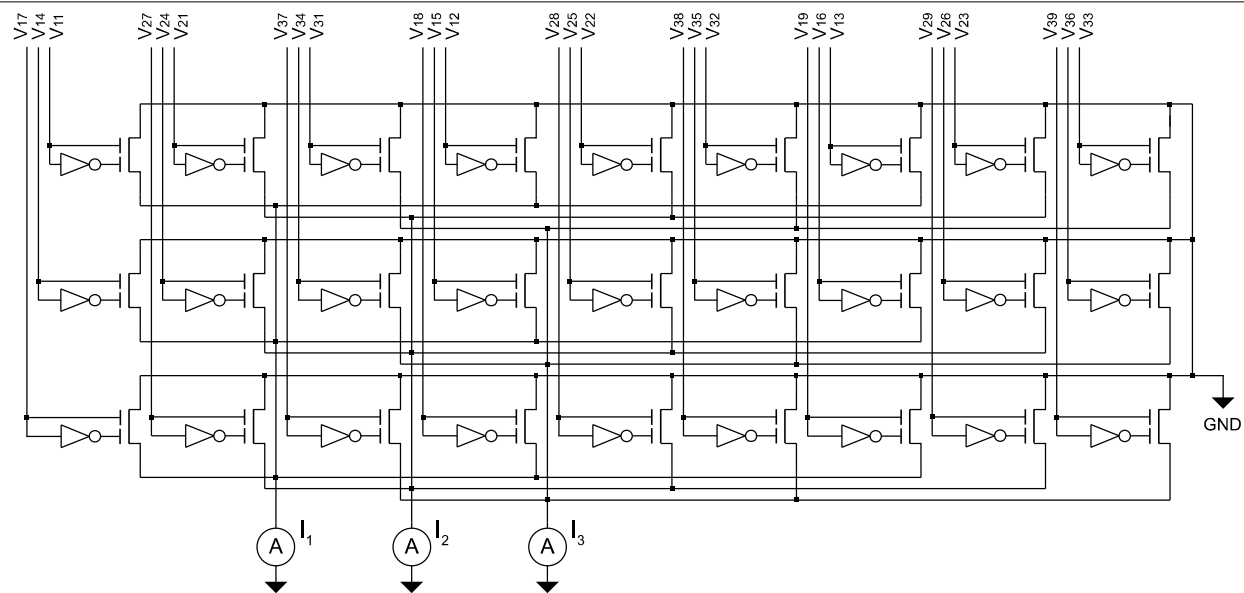
Reprints and permissions information is available at <http://www.nature.com/reprints>.



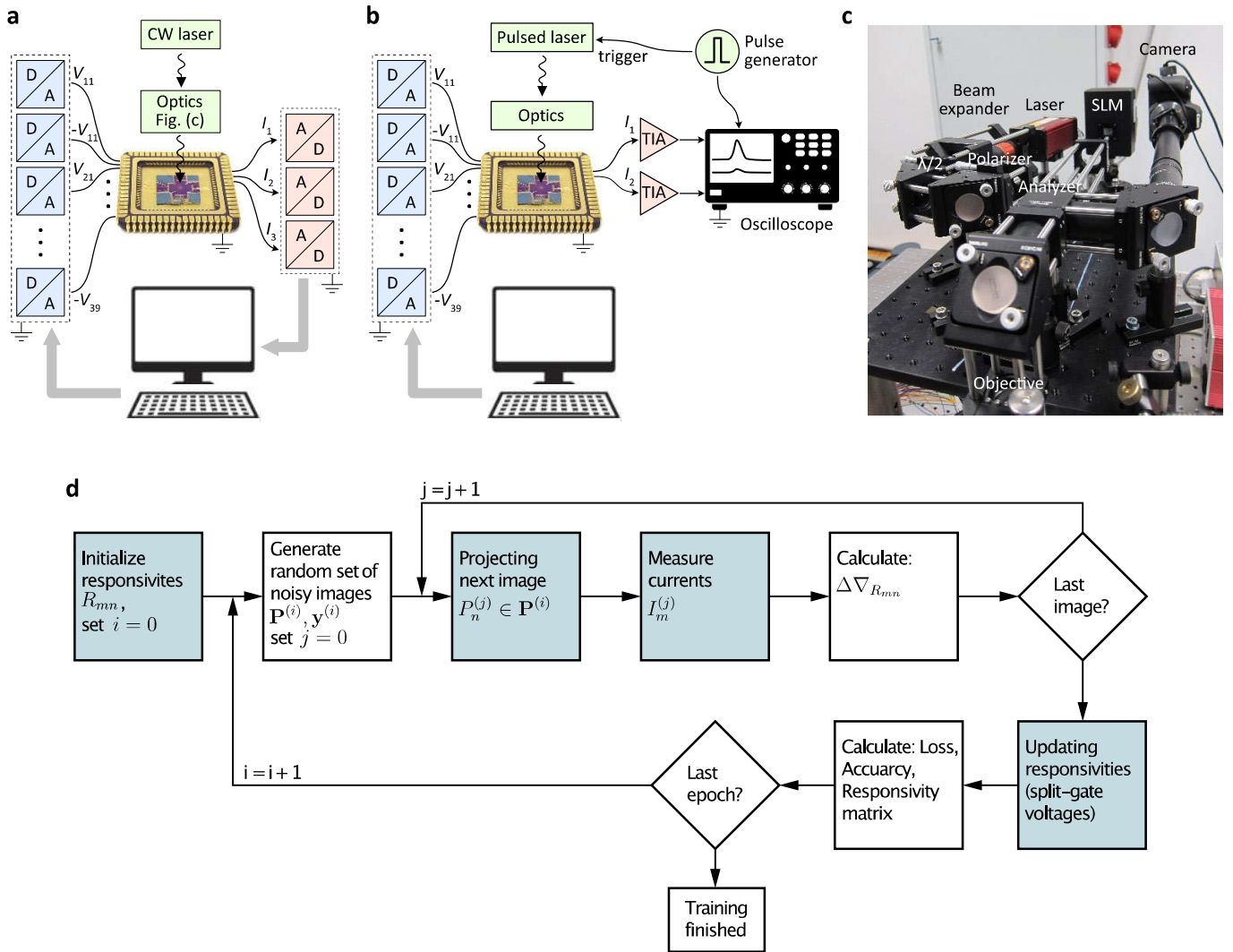
Extended Data Fig. 1 | Photodiode array uniformity. Gate tunability of the responsivities of all 27 photodetectors. One of the detector elements (pixel 7, subpixel 2) did not show any response to light (due to a broken electrical wire), which, however, had no crucial influence on the overall system performance.



Extended Data Fig. 2 | Photodiode characteristics. **a**, Current-voltage characteristic curve under dark (blue) and illuminated (green) conditions. The series resistance R_s and shunt resistance R_{sh} are $10^6 \Omega$ and $10^9 \Omega$, respectively. For zero-bias operation, we estimate a noise-equivalent power of $NEP = I_{th}/R \approx 10^{-13} \text{ W Hz}^{-1/2}$, where $R \approx 60 \text{ mA W}^{-1}$ is the (maximum) responsivity and $I_{th} = \sqrt{4k_B T \Delta f / R_{sh}}$ the thermal noise, where k_B is the Boltzmann constant, Δf is the bandwidth and T is the temperature. **b**, Dependence of the short-circuit photocurrent on the light intensity for different split-gate voltages. Importantly, the response is linear ($I \propto P$), as assumed in equation (1).

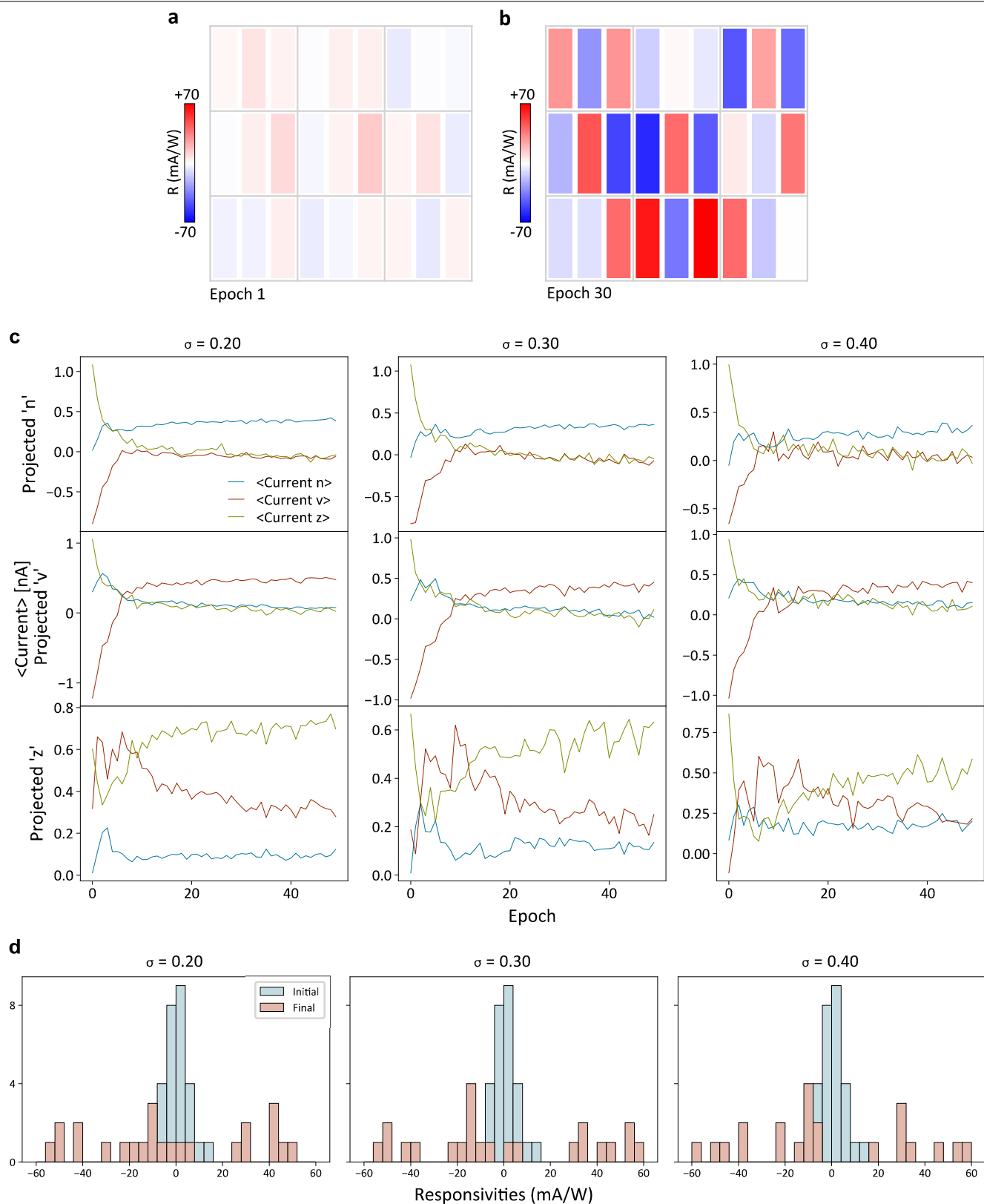


Extended Data Fig. 3 | Circuit of the ANN photodiode array.



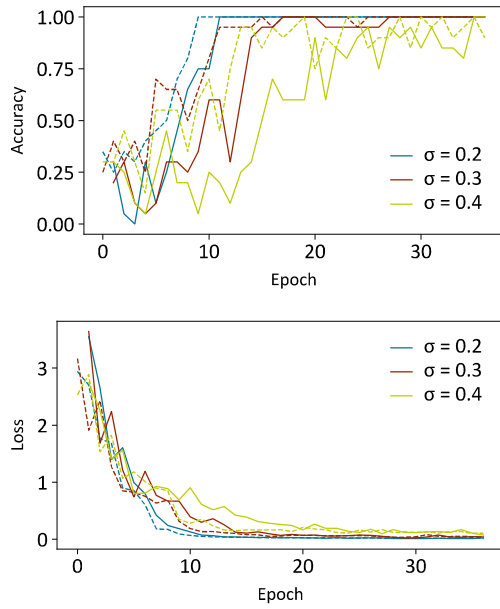
Extended Data Fig. 4 | Experimental setup. **a**, Experimental setup for training the classifier and the autoencoder. CW, continuous wave. **b**, Experimental setup for time-resolved measurements. TIA, transimpedance amplifier. A pulse generator triggers the pulsed laser as well as the oscilloscope. **c**, Photograph of

the optical setup (for schematic see Fig. 2d). **d**, Flow chart of the training algorithm. The blue shaded boxes are interactions with the ANN photodiode array.

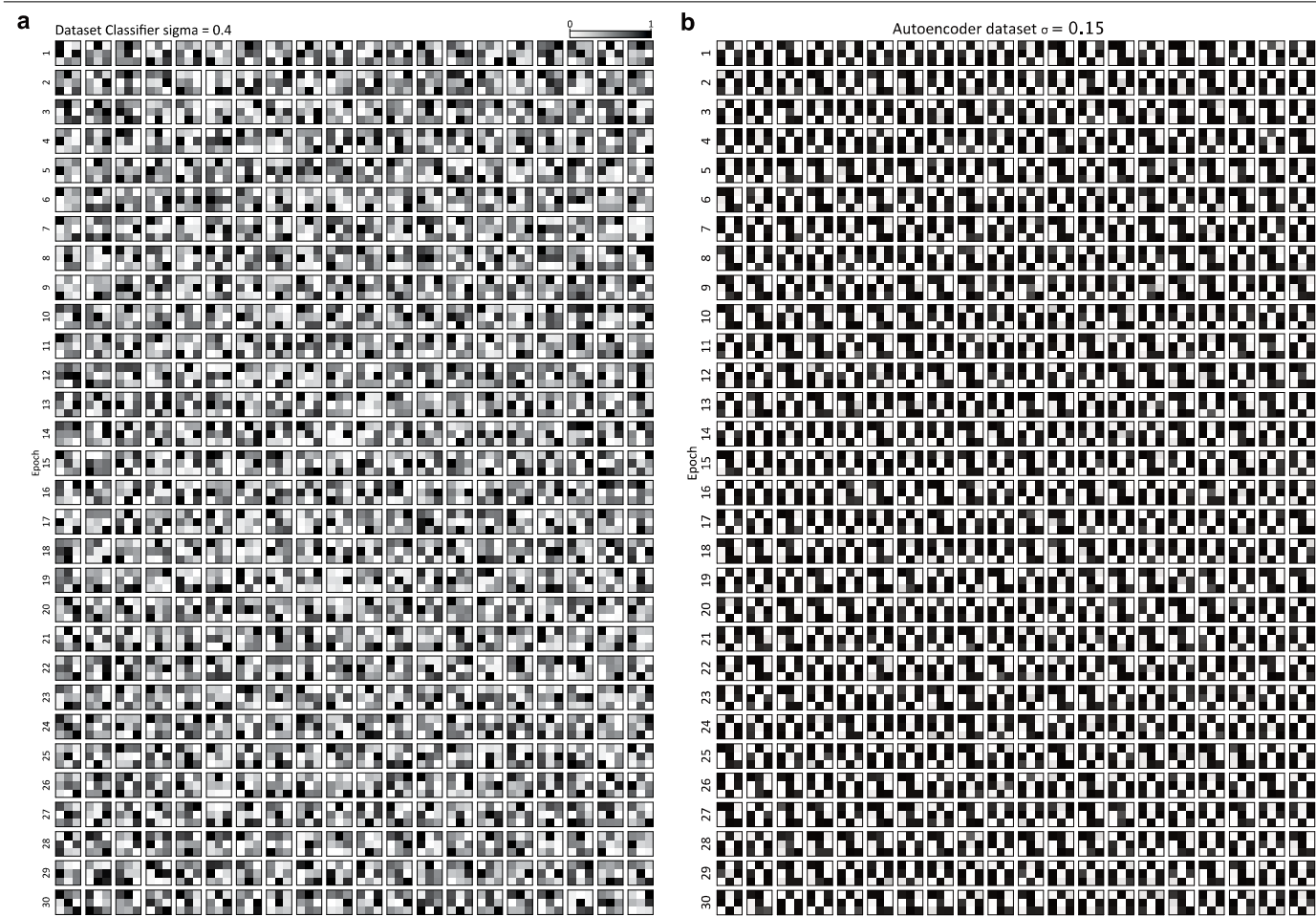


Extended Data Fig. 5 | Classifier training. Photoresponsivity values of all 27 photodetectors with $\sigma=0.3$ training data. **a, b**, Initial (**a**) and epoch 30 (**b**) responsivity values. The weights for the $\sigma=0.2$ and $\sigma=0.4$ training data are

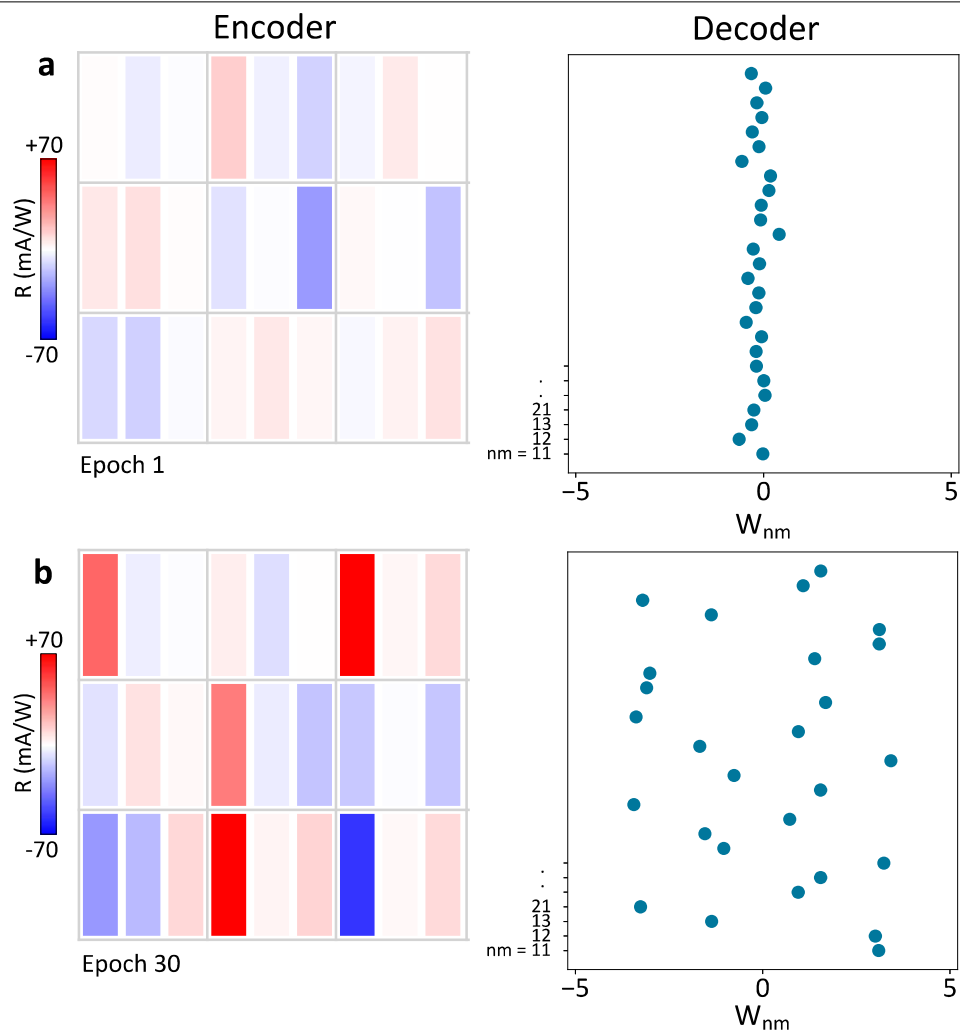
similar. **c**, Measured currents over all epochs for a specific projected letter and at all three noise levels. **d**, Histogram of the initial and final responsivity values for the three different noise levels.



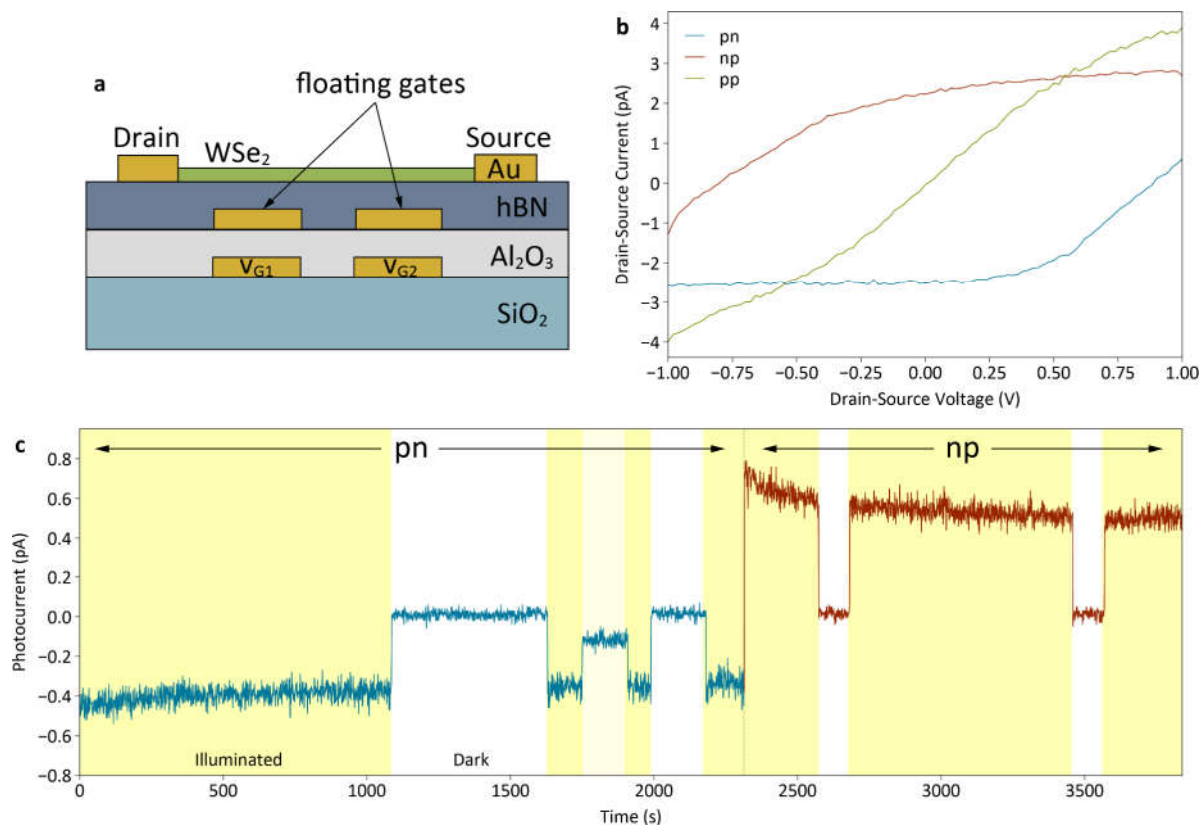
Extended Data Fig. 6 | Comparison with computer simulation. Classifier training of the analogue vision sensor (solid lines) and simulation of the system on a computer (dashed lines) for different data noise levels σ . The same ANN architecture, input data, effective learning rate and starting weights have been used. The same accuracy and loss are eventually reached after training. The slightly slower convergence of the analogue implementation compared with the simulation reflects the nonidealities (defective subpixel, device-to-device variations) of the former. Further discussion on the impact of nonidealities is provided in Extended Data Fig. 10.



Extended Data Fig. 7 | Training datasets. a, b, Dataset of 30 epochs of classifier (a) and autoencoder training (b) with a test data noise level of $\sigma=0.4$ and $\sigma=0.15$ respectively.



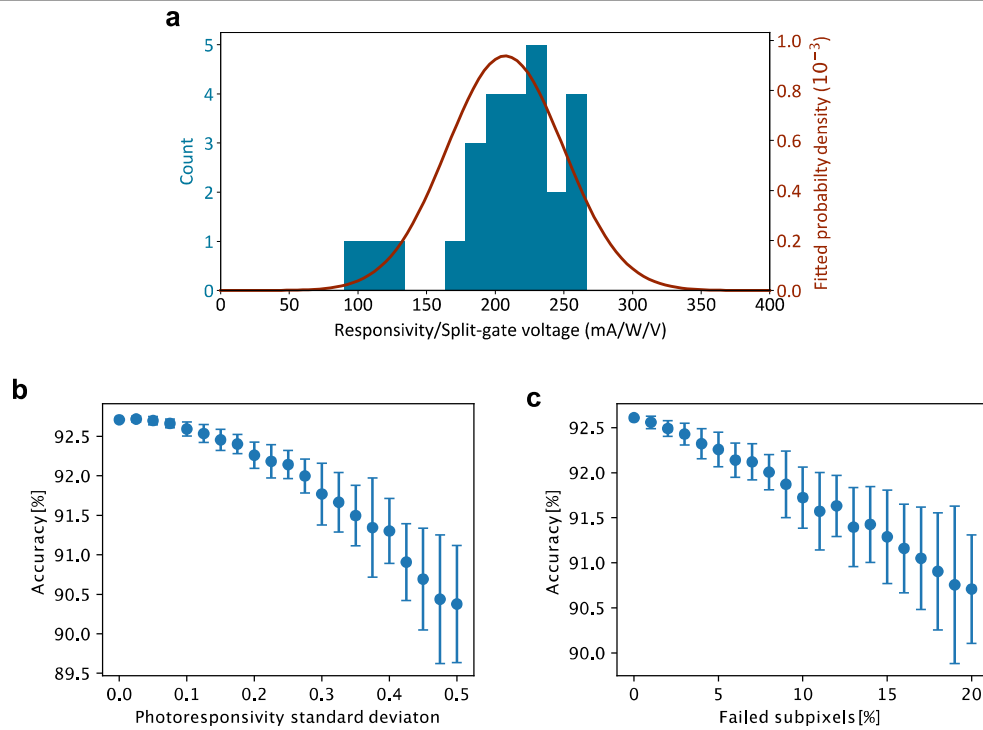
Extended Data Fig. 8 | Autoencoder photoresponsivities/weights. a, b, Initial (a) and epoch 30 (b) encoder photoresponsivity values (left) and decoder weights (right).



Extended Data Fig. 9 | Floating-split-gate photodiode with memory.

a, Schematic of the floating gate photodiode. The addition of 2-nm-thick Au layers, sandwiched between Al₂O₃ and hexagonal boron nitride (hBN), enables the storage of electric charge when a gate voltage is applied to the device, acting as a floating-gate memory. **b**, Electronic characteristic curves of the photodiode operated in p-n, n-p and p-p configurations. **c**, The ability of the device to 'remember' the previous configuration can be verified from the time-resolved photocurrent measurement. The measurement is performed as follows: the back-gate voltages are set to V_{G1} = +5 V and V_{G2} = -5 V and are then disconnected, that is, there is no longer an applied gate voltage and the only

electric field is that generated by the charge stored on the floating electrodes. The short-circuit photocurrent is then measured upon optical illumination. The light is then switched off, at -1,100 s, with a corresponding drop of the photocurrent to zero. After -1,600 s, the light is switched on again, causing the current to reach its initial value, and then a smaller value when the intensity of the light is reduced (-1,700 s). After -2,300 s, the opposite voltage configuration is applied to the back gates (V_{G1} = -5 V and V_{G2} = +5 V), inducing a polarity inversion that also remains permanent. Now, a positive photocurrent (red line) is obtained.



Extended Data Fig. 10 | Robustness of the network. **a**, Detector uniformity, extracted from Extended Data Fig. 1. The fitted Gaussian probability distribution has a standard deviation of $\sigma = 0.205$ ($40 \text{ mA W}^{-1} \text{ V}^{-1}$). **b**, Monte Carlo simulation of a vision sensor with detector responsivities of a given standard deviation. (The photodetectors of the actual device have a measured photoresponsivity standard deviation of 0.205.) Trained on the MNIST

database of handwritten digits, the classifier has 784 pixels and 10 subpixels per pixel. For each data point, 50 random photoresponsivity variations were evaluated. **c**, Accuracy dependence on the number of (randomly chosen) defective subpixels. The same ANN and Monte Carlo simulation scheme as in **b** were used. For each data point, 50 random sets of modified photoresponsivities were evaluated.

High-pressure strengthening in ultrafine-grained metals

<https://doi.org/10.1038/s41586-020-2036-z>

Received: 20 January 2019

Accepted: 3 December 2019

Published online: 24 February 2020

 Check for updates

Xiaoling Zhou^{1,2,3,14}, Zongqiang Feng^{4,14}, Linli Zhu^{5,6,14}, Jianing Xu^{1,7,14}, Lowell Miyagi³, Hongliang Dong¹, Hongwei Sheng¹, Yanju Wang¹, Quan Li^{8,9,10}, Yanming Ma^{8,9,10}, Hengzhong Zhang¹, Jinyuan Yan², Nobumichi Tamura², Martin Kunz², Katie Lutker¹¹, Tianlin Huang¹², Darcy A. Hughes¹³, Xiaoxu Huang⁴✉ & Bin Chen¹✉

The Hall–Petch relationship, according to which the strength of a metal increases as the grain size decreases, has been reported to break down at a critical grain size of around 10 to 15 nanometres^{1,2}. As the grain size decreases beyond this point, the dominant mechanism of deformation switches from a dislocation-mediated process to grain boundary sliding, leading to material softening. In one previous approach, stabilization of grain boundaries through relaxation and molybdenum segregation was used to prevent this softening effect in nickel–molybdenum alloys with grain sizes below 10 nanometres³. Here we track in situ the yield stress and deformation texturing of pure nickel samples of various average grain sizes using a diamond anvil cell coupled with radial X-ray diffraction. Our high-pressure experiments reveal continuous strengthening in samples with grain sizes from 200 nanometres down to 3 nanometres, with the strengthening enhanced (rather than reduced) at grain sizes smaller than 20 nanometres. We achieve a yield strength of approximately 4.2 gigapascals in our 3-nanometre-grain-size samples, ten times stronger than that of a commercial nickel material. A maximum flow stress of 10.2 gigapascals is obtained in nickel of grain size 3 nanometres for the pressure range studied here. We see similar patterns of compression strengthening in gold and palladium samples down to the smallest grain sizes. Simulations and transmission electron microscopy reveal that the high strength observed in nickel of grain size 3 nanometres is caused by the superposition of strengthening mechanisms: both partial and full dislocation hardening plus suppression of grain boundary plasticity. These insights contribute to the ongoing search for ultrastrong metals via materials engineering.

Understanding the strengthening of nanograined metals has been puzzling, as mixed results of both size softening and hardening have been reported^{4–7}. The main challenges in resolving this debate are the difficulty in synthesizing high-quality, ultrafine-grained metal samples for traditional tension or hardness tests and making statistically reproducible measurements. Some researchers have pointed out that reported size softening may be related to the preparation of materials⁸. Porosity, amorphous regions and impurities may be introduced during sample preparation by methods such as inert gas condensation and electrodeposition, leading to softening in micro-hardness measurements and tension tests. Another difficulty is identifying the dominant plastic deformation mechanisms of nanograined metals. Various defects or processes at the nanoscale have been reported,

including dislocations^{9,10}, deformation twinning^{11,12}, stacking faults¹¹, grain boundary (GB) migration¹³, GB sliding¹ and grain rotation^{14,15}. Hence, the processes that dominate plastic deformation and thus determine the strength of nanograined metals are still unclear.

In this study, we use radial diamond anvil cell (DAC) X-ray diffraction (XRD) techniques to track in situ the yield stress and deformation texturing of nickel of various grain sizes. We find that mechanical strengthening can be extended down to a grain size of 3 nm (the smallest we have available), which is much smaller than the previously reported strongest sizes of nanograined metals. This finding pushes mechanical strengthening to the lowest recorded grain size (to our knowledge), demonstrating the potential for achieving ultrahigh strengths in metals.

¹Center for High Pressure Science and Technology Advanced Research, Pudong, Shanghai, China. ²Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

³Department of Geology and Geophysics, University of Utah, Salt Lake City, UT, USA. ⁴International Joint Laboratory for Light Alloys (MOE), College of Materials Science and Engineering, Chongqing University, Chongqing, China. ⁵Center for X-Mechanics, Zhejiang University, Hangzhou, China. ⁶Key Laboratory of Soft Machines and Smart Devices of Zhejiang Province, School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, China. ⁷Department of Physics, Fudan University, Shanghai, China. ⁸State Key Lab of Superhard Materials, College of Physics, Jilin University, Changchun, China. ⁹International Center for Computational Method and Software, College of Physics, Jilin University, Changchun, China. ¹⁰International Center for Future Science, Jilin University, Changchun, China. ¹¹Department of Chemistry, University of California, Berkeley, CA, USA. ¹²Shenyang National Laboratory for Materials Science, Chongqing University, Chongqing, China. ¹³Unaffiliated, Fremont, CA, USA. ¹⁴These authors contributed equally: Xiaoling Zhou, Zongqiang Feng, Linli Zhu, Jianing Xu. ✉e-mail: xiaoxu.huang@cqu.edu.cn; chenbin@hpstar.ac.cn

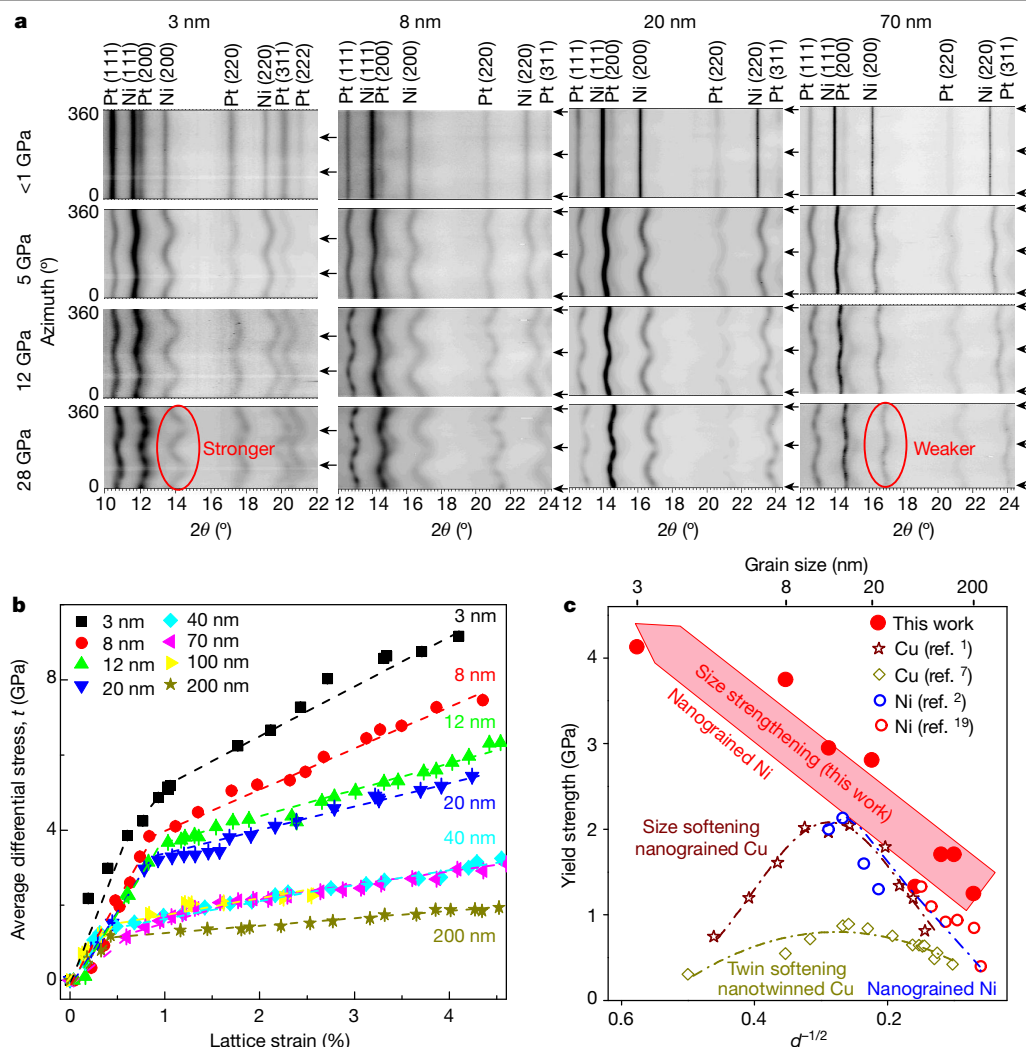


Fig. 1 | Size strengthening of nanogained nickel. **a**, Azimuthally (0–360°) unrolled diffraction images of nickel at different pressures. The black arrows indicate the axial compression direction. Each measurement is repeated at least two times. **b**, Differential stress versus the lattice strain of nickel (see Supplementary Information). We note that for some of the data points the error bars (standard deviations; see equations (6)–(9) and Supplementary Information) are smaller than the sizes of symbols. **c**, Extrapolated yield strength of nickel at ambient conditions without GB sliding (from EVPSC

simulations) versus grain size. The yield strength of nanogained Cu is obtained from molecular dynamics simulations (ref. 1) and experimental data of nanotwinned Cu (ref. 7) and nanogained Ni (refs. 2,19). The yield strength value of nickel in ref. 2 is taken as one-third of its hardness. For nanotwinned Cu, d represents the twin thickness. The inverse Hall–Petch effect has been reported for both Cu (refs. 1,7) and Ni (ref. 2). The smallest grain size of nickel in the study of ref. 2 is 12 nm.

Radial DAC XRD experiments (Extended Data Fig. 1, see Supplementary Information) were performed at beamline 12.2.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory, and at the Shanghai Synchrotron Radiation Facility. Eight nickel samples with particle (grain) sizes ranging from 3 nm to 200 nm (Extended Data Figs. 2, 3) were measured. The relatively narrow size distributions allow for the investigation of the size dependence of the material's strength. In a sample under uniaxial compression, the stress can be separated into hydrostatic and deviatoric stress components. The differential stress between the maximum and minimum compression directions can be obtained using deviatoric strain theory¹⁶ (see Supplementary Information). The measured XRD peak positions provide information on differential strains as well as differential stresses (Fig. 1a). Plastic deformation has an influence on deviatoric strains measured using diffraction, and so the differential strain/stress measured with radial DAC XRD can capture the transition from elastic- to plastic-deformation-dominant behaviours and can provide information on yield strength, strain hardening and so on (Fig. 1 and Extended Data Fig. 4). At the same pressure, the

differential strain of the 3-nm-grain-sized nickel is higher than that of larger-grained counterparts. The larger curvatures (the ellipticity of the XRD rings, which translates into nonlinearity of the lines plotted along the azimuth angle) of diffraction lines for smaller nanocrystals indicate higher elastic deformation and the greater ability of the material to support differential stress in the crystal plane without plastic deformation. We used Rietveld refinement implemented in the MAUD software¹⁷ to analyse the differential strain and texture of our samples at each pressure. The average differential stress of nickel versus its lattice strain can thus be obtained (Fig. 1b) using equations (5) to (9) (see Supplementary Information). To remove the effect of hardening induced by hydrostatic pressure, we performed elasto-viscoplastic self-consistent (EVPSC)¹⁸ simulations (see Supplementary Information) to simulate the stress-strain curves of nickel under ambient conditions (Extended Data Fig. 5). This enables the comparison of our extrapolated strength results at zero pressure with those of conventional tests^{2,19} (Fig. 1c). The stress-strain curves (Fig. 1b and Extended Data Fig. 5c) show that instead of softening, smaller-grained nickel was stronger than its coarser counterparts, in

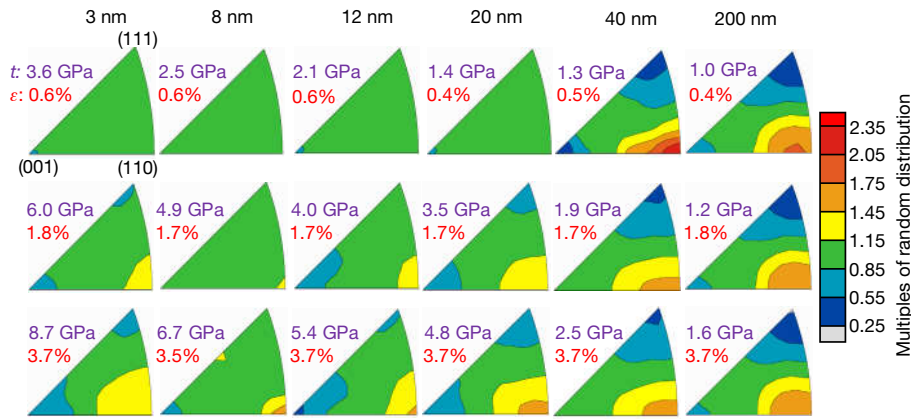


Fig. 2 | Inverse pole figures for the texture evolution of nickel with various grain sizes. Inverse pole figures show the probability of finding the pole to lattice plane in the compression direction. We note that no starting texture exists in all raw (uncompressed powder) samples. τ and ϵ represent differential

stress and lattice strain, respectively. Texture strength is expressed as multiples of random distribution, for which 1 denotes a random distribution and a higher number represents a stronger texture.

strong contrast to the results of previous studies (Fig. 1c). The stress–strain curves of nanograined nickel also show a larger slope/hardening exponent (Fig. 1b), possibly owing to the increased plastic anisotropy in this smaller grain size (Extended Data Fig. 5). We note that a slight strength drop occurs in 40-nm-grain-sized nickel in EVPSC simulations; the cause remains to be further investigated.

The development of in situ deformation textures for nanograined nickel with various grain sizes was captured at different strains. As shown in Fig. 2, nickel samples with larger grain sizes above 20 nm show very strong deformation textures even at low strain. Nanograined nickel samples with grain sizes below 20 nm exhibit very weak deformation textures, indicating that traditional full dislocation activity becomes less active, whereas the strength increases with decreasing grain size. Meanwhile, all of the nickel samples develop a deformation texture, indicating that deformation mechanisms may still be based on dislocation slip and twin formation, since GB-mediated mechanisms would maintain the initial random textures.

Previous simulations^{1,20,21} have suggested that GB deformation plays a decisive part in the deformation mechanisms of sub-10-nm-grain-sized nanomaterials. Those studies proposed that size softening would occur as a result of the transition from dislocation-mediated to GB-mediated mechanisms. In our experiments, however, we observed no size softening but only size strengthening. The uniaxial compressional stress comprises hydrostatic and deviatoric stress components. The shear stress arising from the deviatoric stress could potentially activate GB mechanisms, whereas the hydrostatic stress of the compression increases the critical shear stress for GB migration and sliding, thereby suppressing those mechanisms; see equation (15).

To explore the mechanisms for continuous size strengthening, we simulated the critical stress for activating full and partial dislocations and for activating GB deformation (GB sliding and migration) in nanograined nickel. As shown in Fig. 3a, full dislocations are activated preferentially and are more dominant than partial dislocations above the critical grain size d_c^1 . The dislocation-dominated deformation shifts to GB-dominated deformation for grain sizes below a critical grain size d_c^2 . However, compression has a remarkable effect on this shift. The critical stress for activating GB deformation increases with pressure, resulting in the critical grain size d_c^2 being highly pressure-dependent. For example, the critical grain size for active GB deformation of nickel at >1 GPa is <2 nm (Fig. 3b); this suggests that almost no GB deformation is activated in our experiments because hydrostatic pressure is higher than 1 GPa, that is, the GB-deformation associated softening of nanograins has been greatly inhibited during compression. Consequently, when GB-associated deformation (extrinsic deformation) is

suppressed, the material strength should be determined mainly by intrinsic deformation properties, which are associated with lattice strain and defects in the interiors of grains.

It is known that the critical stress to activate dislocations increases with decreasing grain size. In a simplified analytical dislocation model that considers partial dislocations emitted from GBs of nanograins, the critical stress for emitting a full and partial dislocation²² can be described as:

$$\tau_f = \frac{Gb_f}{d} \quad (1)$$

$$\tau_p = \frac{Gb_p}{3d} + (1 - \delta) \frac{\gamma}{Gb_p} \quad (2)$$

where b_f and b_p are the Burgers vectors of the full and partial dislocations, respectively; G is the shear modulus; γ is the stacking fault energy; and δ is the ratio of equilibrium stacking fault width to grain size. The critical stresses for nucleating both full and partial dislocations increase sharply as the grain size decreases towards the lower limit (Fig. 3a). This leads to the increase of yield strength at small grain size. Furthermore, partial dislocations are preferentially activated and overtake full dislocations below a critical grain size.

We studied the deformation behaviour of nanograined nickel by molecular dynamics simulations²³ (Fig. 3d). Two types of planar defects associated with partial dislocations (that is, nanotwins and stacking faults) as well as full dislocations are found in nanograined nickel under compression. To explore the deformation mechanisms, we conducted transmission electron microscopy (TEM) characterization on the recovered samples. As expected, high densities of full dislocations are seen in the coarse-grained sample (Fig. 4d). Remarkably, full dislocations are prevalent at all average grain sizes including the finest at 3 nm (Fig. 4a–c), although for the 3-nm-grain-sized sample the dislocations were observed in grains with slightly larger sizes than average. A detailed analysis of the four dislocations observed in the lower part of Fig. 4a is shown in the sketch in Fig. 4b based on the Thompson tetrahedron. Each of these dislocations is an extended dislocation composed of a stacking fault and two partial dislocations lying on {111} slip planes. A Lomer–Cottrell lock and a stair rod are formed from the reactions of partial dislocations associated with the upper three dislocations, with the stair rod lying on a {100} plane. These reaction products are immobile and thus provide a strong strengthening effect. A rough estimate of the density of full dislocations based on the dislocations in Fig. 4a suggests a density of about 10^{16} m^{-2} , which provides a strong

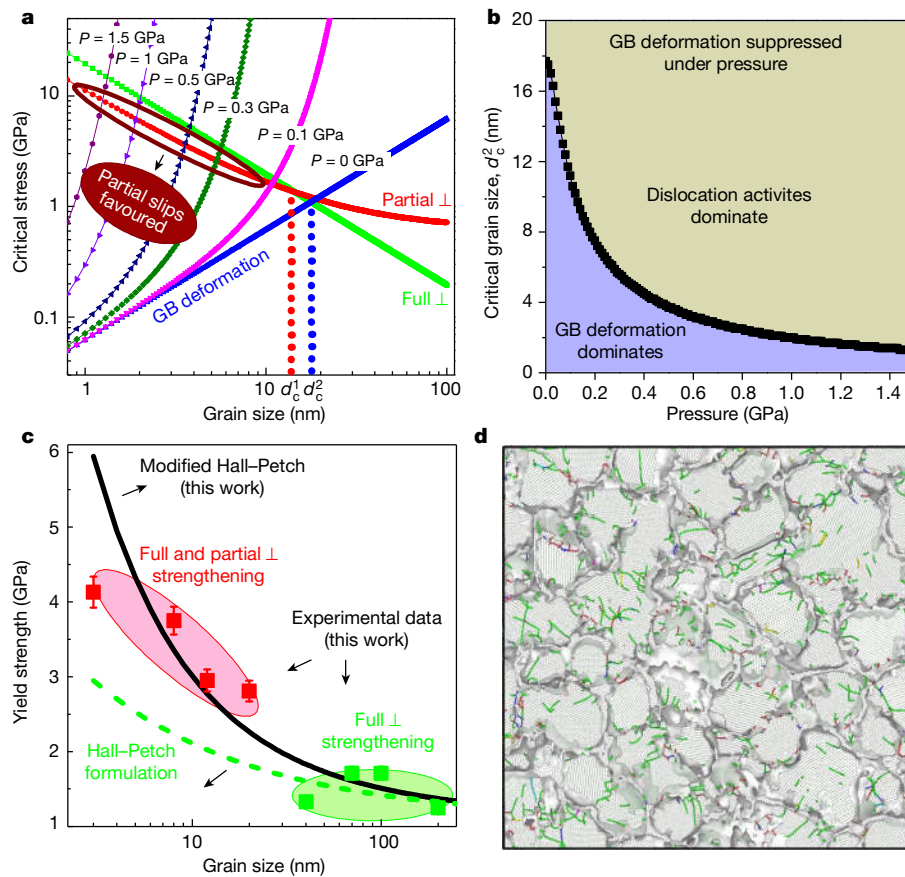


Fig. 3 | Computational simulation results and the modified Hall-Petch relationship. **a**, Comparison of simulated grain-size-dependent critical stresses for activating dislocations and GB deformation in nanograined nickel at different pressures. The \perp symbol represents a dislocation. **b**, Critical grain size d_c^2 as a function of pressure. **c**, The predicted yield strength compared with the experimental data for nanograined nickel. **d**, Classic molecular dynamics

simulation of 3-nm-grain-sized Ni compressed with 10% volume strain. Green indicates partial dislocations associated with stacking faults, twins and grain boundaries. Blue indicates perfect dislocations. Yellow, purple and red indicate a few $1/3 \langle 001 \rangle$ (Hirth), $1/6 \langle 110 \rangle$ (stair-rod) and other types of dislocations.

strengthening component for the flow stress. At the finest grain sizes, nanotwins form bounded by stacking faults, creating important new additions to the deformed structure. These nanotwins further refine the nanostructure and contribute to boundary strengthening by constraining dislocation motion. Steps in the twin boundaries are observed, forming incoherent twin boundaries that contain partial dislocations (Fig. 4). We note that the simultaneous and cooperative activation of different Shockley partial dislocations on parallel and neighbouring glide planes may be responsible for these twins²⁴. Stacking faults may expand under high stress, increasing their energy and making it favourable to form low-energy twins²⁵. Fivefold symmetry twins are also seen in both 3-nm-grain-sized and 20-nm-grain-sized quenched nickel samples (Fig. 4). Fivefold twins may pre-exist in the particles or form by the successive emission of partial dislocations from incoherent twin boundaries with high energy. The non-parallel twin boundaries give rise to strong overlapping of associated lattice strain fields, resulting in higher yield strength compared to those without fivefold twinned structures²⁶. This result is consistent with our observations in the mechanical measurements (Fig. 1b). In short, twinning and stacking faults observed in our TEM measurements originate from the nucleation and motion of partial dislocations. This provides compelling evidence that in the sub-20-nm regime of grain size, full-dislocation-mediated deformation shifts to both full and partial dislocations combined with deformation twinning.

Our strength measurements (Fig. 1b), computational simulations (Fig. 3a, c) and TEM observations (Fig. 4) indicate that a critical grain

size (around 20 nm) exists and corresponds to the shift in deformation mechanisms from full dislocation to full plus partial dislocation mediated deformation. This does not generate a maximum strength at the critical grain size but starts a stronger mode for strengthening. Notably, as shown in Fig. 4, the twins in 20 nm or smaller nickel grains are usually only several nanometres thick, but unlike growth twins, no softening is induced in pressurized nickel nanograins. Instead, size strengthening of nickel is even more pronounced in the smaller size range of nanograins. As shown in Fig. 3c, for grain sizes below 20 nm the measured yield strength of nanograined nickel largely deviates from the trend predicted by the traditional Hall-Petch model. Considering that the contribution of partial dislocations becomes important in fine nanograins, we propose a modified Hall-Petch relationship as follows:

$$\sigma_y = \sigma_0 + \frac{k_0}{\sqrt{d}} + \frac{k_1}{d} \quad (3)$$

where σ_y and d represent the yield strength and grain size, respectively, and σ_0 , k_0 and k_1 are constants. The first two terms represent the friction stress and Hall-Petch formulation associated with full dislocation boundary interaction. The third term is related to the partial dislocation contribution to yielding, which is inversely proportional to grain size d according to equation (2). The fitting of our experimental data with equation (3) shows that this new model reflects the effects of both full and partial dislocations, and can describe the size strengthening of

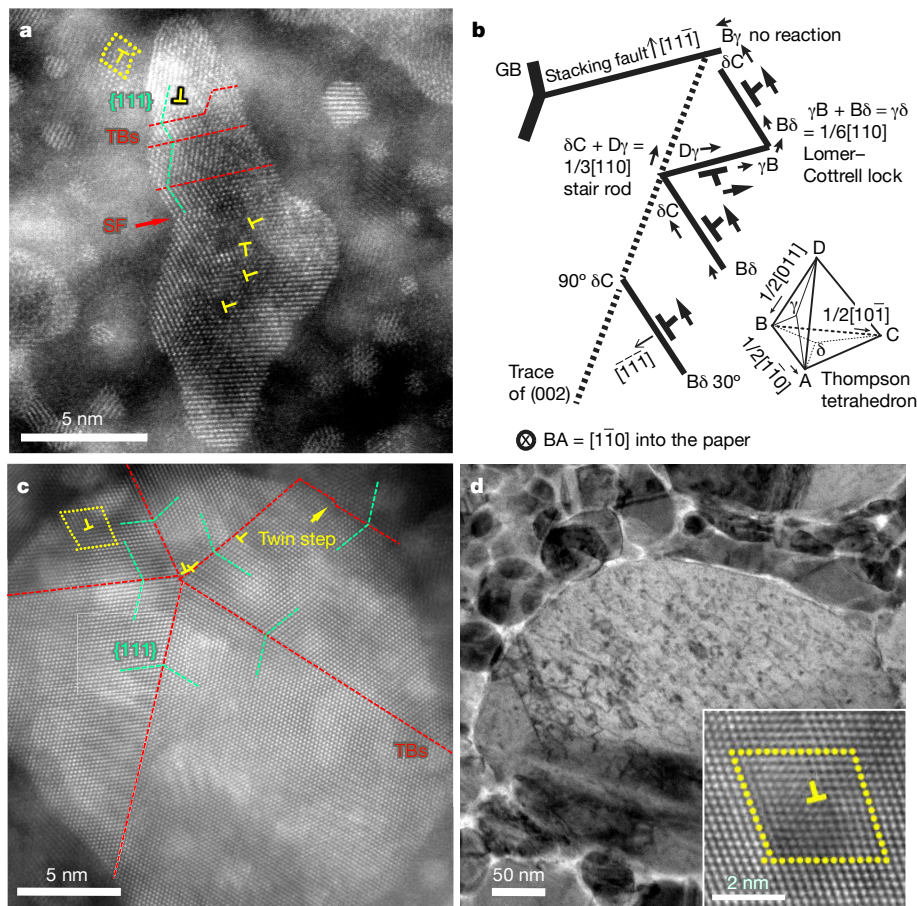


Fig. 4 | TEM examinations of nickel samples quenched from 40 GPa of three grain sizes. a, 3 nm; c, 20 nm; and d, 200 nm. Panel b is a sketch showing the analysis of dislocations observed in the lower part of the middle grain in panel a. We note the reactions of partial dislocations forming Lomer–Cottrell locks

metals over a wide size range. We note that this fit gives a high friction stress of about 1.1 GPa, k_1 of 9 MPa μm and a low k_0 of 101 MPa $\mu\text{m}^{1/2}$ compared to conventionally deformed Ni with 20 MPa and 158 MPa $\mu\text{m}^{1/2}$, respectively^{27,28}. To check the generality of the size strengthening for nanograined metals, we conducted similar high-pressure deformation experiments on nanograined gold and palladium. A similar enhanced strengthening effect at the lowest grain sizes was observed, which indicates that this full plus partial dislocation-mediated strengthening is common in compressed nanograined metals.

This size strengthening effect may apply not only to high-pressure cases but also provide guidance for applications at ambient conditions. A recent study³ reported that a twofold increase in hardness was achieved in nanograined Ni–Mo alloys by stabilizing GBs through Mo segregation. By using this technique yield strengths of around 1.6 GPa and 3.8 GPa were achieved in Ni and Ni–Mo alloys, respectively. In our experiments, an ultrahigh strength of about 4.2 GPa is achieved in pure nickel grains. This result suggests that compression is an effective method of suppressing GB sliding and migration in order to achieve ultrahigh strength.

This is also supported by the observation that the measured strength of coarse nickel in our compression test is higher than in conventional tension tests^{2,19}. In real applications, materials could be under either tension or compression. Tension tests are common in traditional mechanical characterization. However, evaluation of strength by tensile loading is often technically difficult for nanograined metals especially for sub-10-nm grain sizes. Compressive strength measurements using radial DAC XRD enables study of the mechanical properties of even

and a stair rod. Stacking faults (SFs), twin boundaries (TBs) and a few full dislocations can be found in 3 and 20 nm nickel samples. A high density of full dislocations is observed in the 200 nm nickel grains. The inset to **d** shows a high-resolution image of a full dislocation.

sub-10-nm-sized metals. In this synchrotron-based study, deformation behaviour and yield strength are obtained from the lattice changes of a large quantity of nanograins and exhibit reproducible trends in strength and grain size. Additionally, extrinsic factors like impurities and amorphous regions that may be introduced during conventional sample preparation could strongly affect the mechanical behaviour of nanograined metals. In our method, the strength in pure nickel grains is determined by the internal piezometer of crystalline lattice strain, which mitigates the effects of extrinsic factors.

Experimental work also indicates that partial-dislocation-associated mechanisms improve the thermal stability of grain boundaries of nanograins²⁹. If the grain boundaries of nanograined metals are sintered without grain coarsening, for example, through severe plastic deformation or explosive shock annealing³⁰, large pieces of nanograined metals with ultrahigh strength could potentially be fabricated for mass applications. In summary, achieving an ultrahigh strength in pure nickel through grain refinement and suppression of GB plasticity provides a new strategy for designing ultrastrong, ultrahard metals for future applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2036-z>.

1. Schiøtz, J. & Jacobsen, K. W. A maximum in the strength of nanocrystalline copper. *Science* **301**, 1357–1359 (2003).
2. Schuh, C., Nieh, T. & Yamasaki, T. Hall–Petch breakdown manifested in abrasive wear resistance of nanocrystalline nickel. *Scr. Mater.* **46**, 735–740 (2002).
3. Hu, J., Shi, Y. N., Sauvage, X., Sha, G. & Lu, K. Grain boundary stability governs hardening and softening in extremely fine nanograined metals. *Science* **355**, 1292–1296 (2017).
4. Knapp, J. & Follstaedt, D. Hall–Petch relationship in pulsed-laser deposited nickel films. *J. Mater. Res.* **19**, 218–227 (2004).
5. Meyers, M. A., Mishra, A. & Benson, D. J. Mechanical properties of nanocrystalline materials. *Prog. Mater. Sci.* **51**, 427–556 (2006).
6. Huang, X., Hansen, N. & Tsuji, N. Hardening by annealing and softening by deformation in nanostructured metals. *Science* **312**, 249–251 (2006).
7. Lu, L., Chen, X., Huang, X. & Lu, K. Revealing the maximum strength in nanotwinned copper. *Science* **323**, 607–610 (2009).
8. Koch, C. C. & Narayan, J. The Inverse Hall–Petch Effect—Fact or Artifact? *MRS Online Proc. Lib. Arch.* **634**, B5.1.1, <https://www.cambridge.org/core/journals/mrs-online-proceedings-library-archive/article/inverse-hall-etch-effect-fact-or-artifact/E7759F8A3266F51367E3A87EFE13FA2B> (2000).
9. Chen, B. et al. Texture of nanocrystalline nickel: probing the lower size limit of dislocation activity. *Science* **338**, 1448–1451 (2012).
10. Hughes, D. & Hansen, N. Exploring the limit of dislocation based plasticity in nanostructured metals. *Phys. Rev. Lett.* **112**, 135504 (2014).
11. Chen, M. et al. Deformation twinning in nanocrystalline aluminum. *Science* **300**, 1275–1277 (2003).
12. Yamakov, V., Wolf, D., Phillpot, S. R., Mukherjee, A. K. & Gleiter, H. Dislocation processes in the deformation of nanocrystalline aluminium by molecular-dynamics simulation. *Nat. Mater.* **1**, 45–49 (2002).
13. Shan, Z. et al. Grain boundary-mediated plasticity in nanocrystalline nickel. *Science* **305**, 654–657 (2004).
14. Zhou, X. et al. Reversal in the size dependence of grain rotation. *Phys. Rev. Lett.* **118**, 096101 (2017).
15. Chen, B., Zhu, L., Xin, Y. & Lei, J. Grain rotation in plastic deformation. *Quantum Beam Sci.* **3**, 17 (2019).
16. Singh, A. K., Balasingh, C., Mao, H.-k., Hemley, R. J. & Shu, J. Analysis of lattice strains measured under nonhydrostatic pressure. *J. Appl. Phys.* **83**, 7567 (1998).
17. Lutterotti, L., Vasin, R. & Wenk, H.-R. Rietveld texture analysis from synchrotron diffraction images. I. Calibration and basic analysis. *Powder Diff.* **29**, 76–84 (2014).
18. Wang, H., Wu, P., Tomé, C. & Huang, Y. A finite strain elastic–viscoplastic self-consistent model for polycrystalline materials. *J. Mech. Phys. Solids* **58**, 594–612 (2010).
19. Ebrahimi, F., Bourne, G., Kelly, M. S. & Matthews, T. Mechanical properties of nanocrystalline nickel produced by electrodeposition. *Nanostruct. Mater.* **11**, 343–350 (1999).
20. Yamakov, V., Wolf, D., Phillpot, S., Mukherjee, A. & Gleiter, H. Deformation-mechanism map for nanocrystalline metals by molecular-dynamics simulation. *Nat. Mater.* **3**, 43 (2004).
21. Van Swygenhoven, H., Derlet, P. & Frøseth, A. Stacking fault energies and slip in nanocrystalline metals. *Nat. Mater.* **3**, 399–403 (2004).
22. Zhu, Y., Liao, X. & Wu, X. Deformation twinning in nanocrystalline materials. *Prog. Mater. Sci.* **57**, 1–62 (2012).
23. Zepeda-Ruiz, L. A., Stukowski, A., Oppelstrup, T. & Bulatov, V. V. Probing the limits of metal plasticity with molecular dynamics simulations. *Nature* **550**, 492–495 (2017).
24. Li, B., Li, B., Wang, Y., Sui, M. & Ma, E. Twinning mechanism via synchronized activation of partial dislocations in face-centered-cubic materials. *Scr. Mater.* **64**, 852–855 (2011).
25. Wang, J. & Huang, H. Shockley partial dislocations to twin: another formation mechanism and generic driving force. *Appl. Phys. Lett.* **85**, 5983–5985 (2004).
26. Zhang, Z., Huang, S., Chen, L., Zhu, Z. & Guo, D. Formation mechanism of fivefold deformation twins in a face-centered cubic alloy. *Sci. Rep.* **7**, 45405 (2017).
27. Thompson, A. A. W. Yielding in nickel as a function of grain or cell size. *Acta Metall.* **23**, 1337–1342 (1975).
28. Hughes, D. & Hansen, N. The microstructural origin of work hardening stages. *Acta Mater.* **148**, 374–383 (2018).
29. Huang, Q. et al. Nanotwinned diamond with unprecedented hardness and stability. *Nature* **510**, 250–253 (2014).
30. Zheng, S. et al. High-strength and thermally stable bulk nanolayered composites due to twin-induced interfaces. *Nat. Commun.* **4**, 1696 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgements We thank K. Lu, D. J. Jensen, N. Hansen, S.-I. Karato, G. Fan, J. Liu and F. Zhao for pre-review and discussions. X. Z. thanks F. Lin for EVPSC tutoring. We acknowledge support from the National Natural Science Foundation of China (NSFC) under grant numbers 11621062, 11772294, U1530402 and 11811530001. X.Z. acknowledges the Advanced Light Source Doctoral Fellowship in Residence Program and Collaborative Postdoctoral Fellowship Program. L.Z. acknowledges support from the Fundamental Research Funds for the Central Universities of China (2018XZZX001–05). L.M. acknowledges support from CDAC and NSF (EAR-1654687). Z.F., T.H. and X.H. acknowledge support from the National Key Research and Development Program of China (2016YFB0700400). This research used the resources of the Advanced Light Source, which is a DOE Office of Science User Facility under contract number DE-AC02-05CH11231 and the Shanghai Synchrotron Radiation Facility. This research was

partially supported by COMPRES, the Consortium for Materials Properties Research in Earth Sciences under NSF Cooperative Agreement EAR 1606856.

Author contributions B.C. conceived the project. X.H. directed the TEM examinations. J.X., H.D., H.Z. and K.L. performed the nanocrystal synthesis. X.Z., B.C., L.M., J.Y., N.T. and M.K. performed the high-pressure XRD experiments. Z.F., Y.W., D.A.H., T.H. and X.H. performed the TEM experiments and analysis. L.Z., H.S., Q.L. and Y.M. performed the computational and molecular dynamics simulations. X.Z. and L.M. performed the EVPSC modelling. X.Z. and B.C. wrote the manuscript. All authors discussed the results and commented on the manuscript.

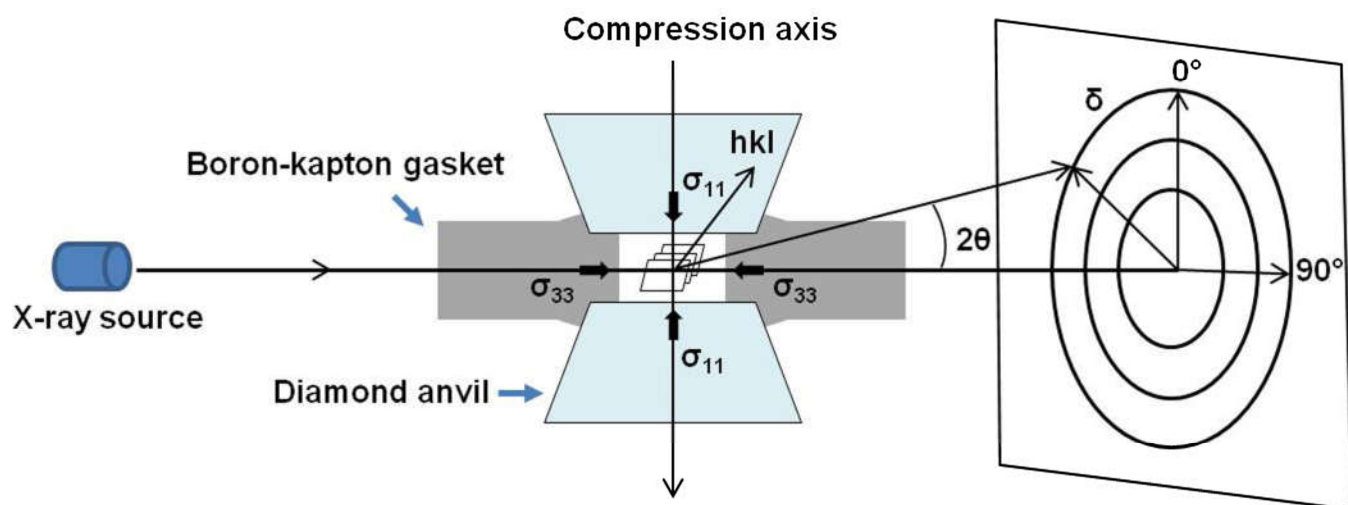
Competing interests The authors declare no competing interests.

Additional information

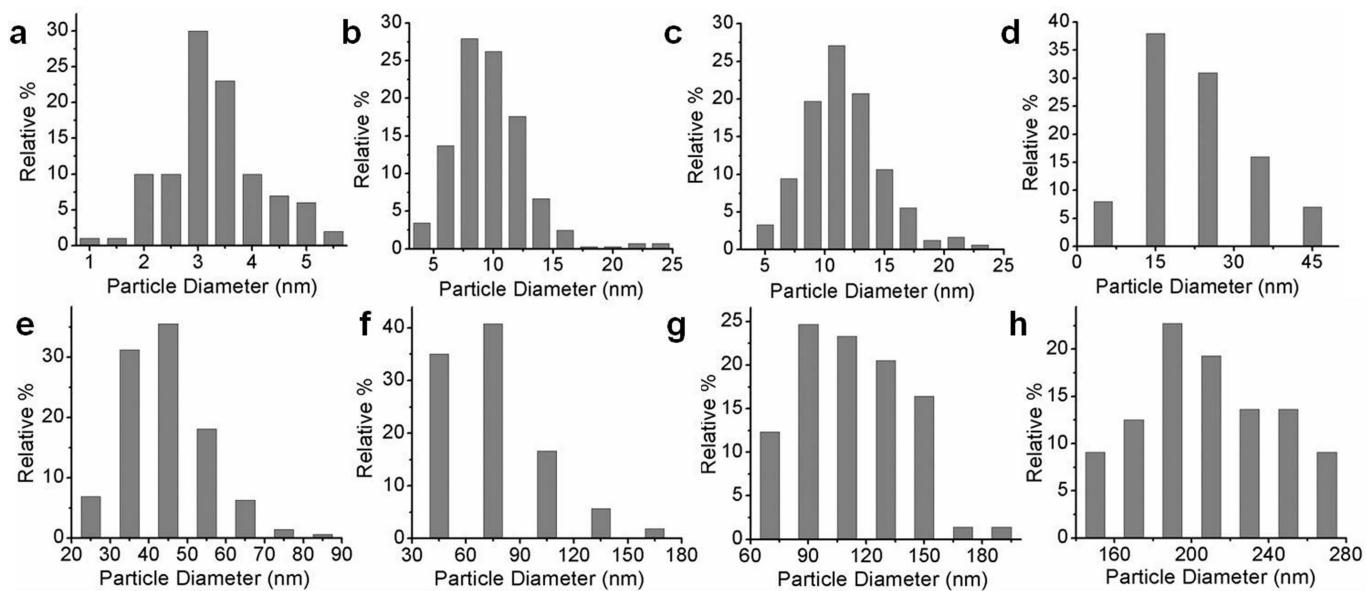
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2036-z>.

Correspondence and requests for materials should be addressed to X.H. or B.C.

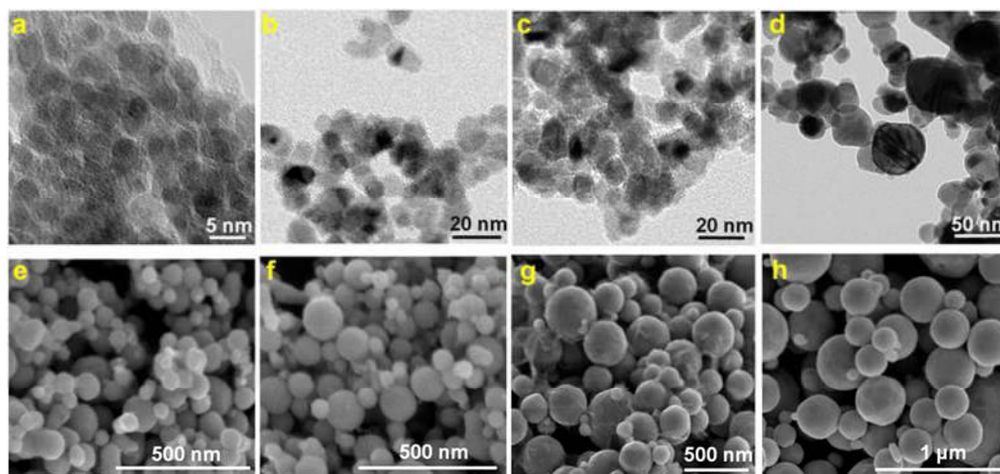
Reprints and permissions information is available at <http://www.nature.com/reprints>.



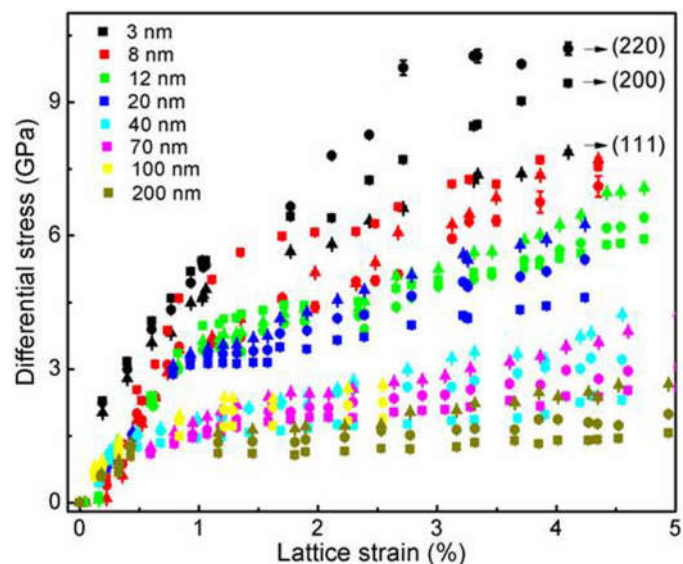
Extended Data Fig. 1 | The experimental setup of radial DAC XRD. Kapton is a polyimide film. hkl represents the lattice planes; δ is the azimuthal angle; and θ represents the diffraction angle.



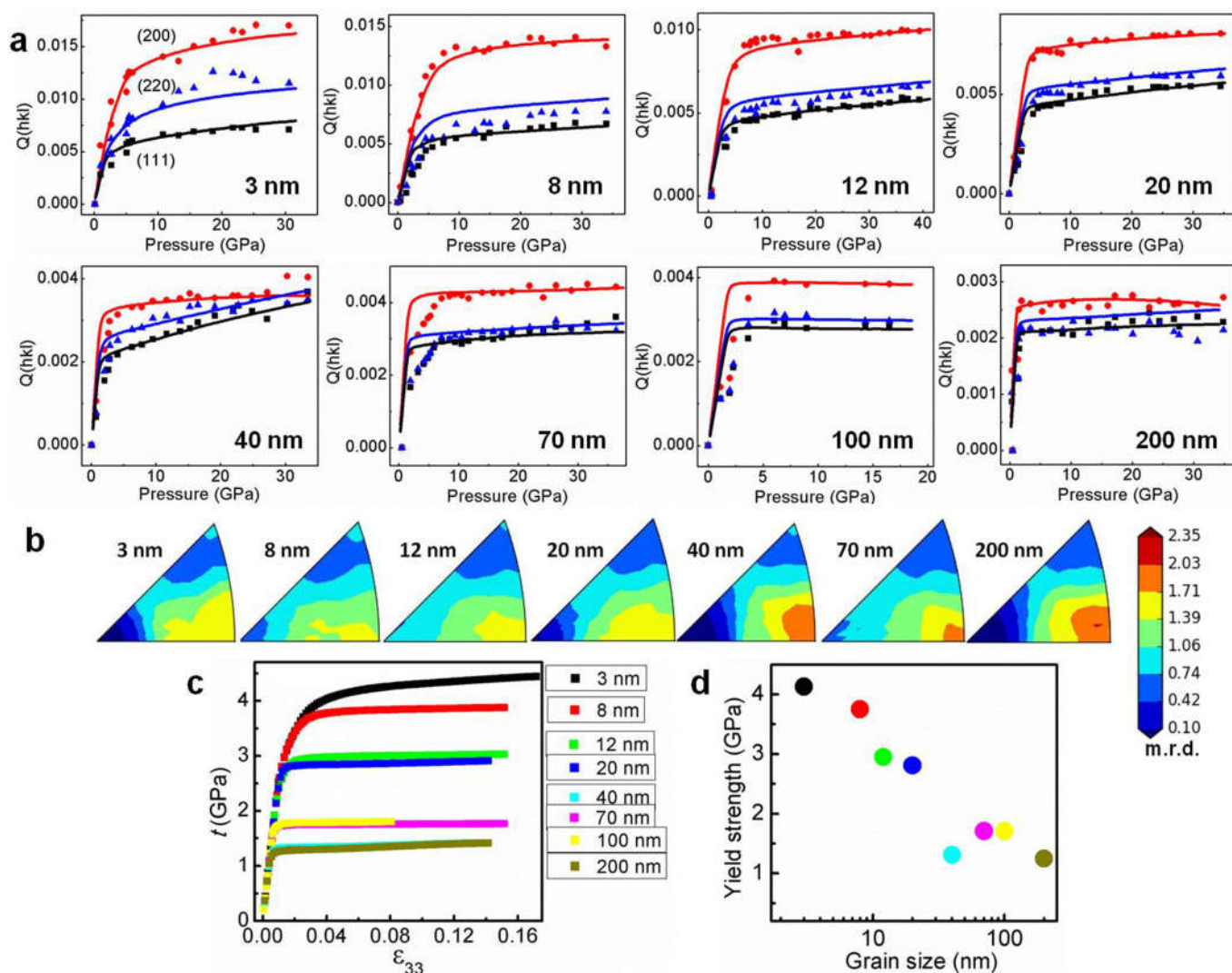
Extended Data Fig. 2 | Grain size distribution of nickel samples. **a–d**, Grain size distributions in 3-nm, 8-nm, 12-nm and 20-nm nickel; **e–h**, Grain size distribution of 40-nm, 70-nm, 100-nm and 200-nm nickel. The particle sizes of the nickel samples were re-checked with XRD characterization.



Extended Data Fig. 3 | Raw powder samples. **a–d**, TEM images of raw powder samples of 3 nm (**a**), 8 nm (**b**), 12 nm (**c**) and 20 nm (**d**) nickel powder before compression. **e–h**, Scanning electron microscopy characterization of 40 nm (**e**), 70 nm (**f**), 100 nm (**g**) and 200 nm (**h**) nickel powder before compression.



Extended Data Fig. 4 | Plot of differential stress versus hydrostatic lattice strain in the nickel of various grain sizes. The circles, squares and triangles represent (220), (200) and (111) lattice planes, respectively. Strong strength anisotropy is exhibited for different lattice planes, especially at smaller grain sizes. The lattice strain is calculated from the relative change in the unit cell parameter at a given applied stress to the unit cell parameter under ambient pressure (see Supplementary Information). The error bars for differential stress is calculated based on the error of deviatoric strain $Q(hkl)$ and equations (6) to (9). Note that for some of the data points the error bars (see Supplementary Information for definition) are smaller than the sizes of symbols.



Extended Data Fig. 5 | EVPSC modelling results of nickel. **a**, Comparison between simulated $Q(hkl)$ curves versus pressure and measured $Q(hkl)$ values (solid symbols) obtained from experiments. **b**, Simulated texture of nickel at the highest strain (pressure). **c**, Simulated differential stress of nickel versus

plastic strain during zero pressure compression. **d**, Extrapolated yield strength of nickel at ambient conditions without GB sliding. The size-strengthening trend is consistent with that shown in Fig. 1b, although the strength of 40-nm-grain-size nickel obtained with EVPSC is slightly lower.

Enhancing crystal growth using polyelectrolyte solutions and shear flow

<https://doi.org/10.1038/s41586-020-2042-1>

Received: 3 April 2019

Accepted: 2 December 2019

Published online: 4 March 2020

 Check for updates

Jian-Ke Sun^{1,3,6}, Yaroslav I. Sobolev^{1,6}, Weiye Zhang⁴, Qiang Zhuang^{1,5} & Bartosz A. Grzybowski^{1,2}✉

The ability to grow properly sized and good quality crystals is one of the cornerstones of single-crystal diffraction, is advantageous in many industrial-scale chemical processes^{1–3}, and is important for obtaining institutional approvals of new drugs for which high-quality crystallographic data are required^{4–7}. Typically, single crystals suitable for such processes and analyses are grown for hours to days during which any mechanical disturbances—believed to be detrimental to the process—are carefully avoided. In particular, stirring and shear flows are known to cause secondary nucleation, which decreases the final size of the crystals (though shear can also increase their quantity^{8–14}). Here we demonstrate that in the presence of polymers (preferably, polyionic liquids), crystals of various types grow in common solvents, at constant temperature, much bigger and much faster when stirred, rather than kept still. This conclusion is based on the study of approximately 20 diverse organic molecules, inorganic salts, metal–organic complexes, and even some proteins. On typical timescales of a few to tens of minutes, these molecules grow into regularly faceted crystals that are always larger (with longest linear dimension about 16 times larger) than those obtained in control experiments of the same duration but without stirring or without polymers. We attribute this enhancement to two synergistic effects. First, under shear, the polymers and their aggregates disentangle, compete for solvent molecules and thus effectively ‘salt out’ (that is, induce precipitation by decreasing solubility of) the crystallizing species. Second, the local shear rate is dependent on particle size, ultimately promoting the growth of larger crystals (but not via surface-energy effects as in classical Ostwald ripening). This closed-system, constant-temperature crystallization driven by shear could be a valuable addition to the repertoire of crystal growth techniques, enabling accelerated growth of crystals required by the materials and pharmaceutical industries.

Although the phenomena we describe are observed even in solutions stirred by an ordinary magnetic stir bar (Supplementary Video 1), most experiments were performed in a standardized Couette cell, with a gap of $d = 1$ mm and the inner cylinder ($r_i = 4$ mm in radius) rotating at a constant angular velocity, usually $\omega = 400$ rpm, corresponding to shear rate $\dot{\gamma} = 167 \text{ s}^{-1}$ (Fig. 1) but down to 60 rpm in some control experiments (see Fig. 4a). In all experiments, the Reynolds number (for the inner cylinder) $Re = r_i \omega d / \nu$ (where ν is kinematic viscosity) was smaller than 2, ensuring simple Couette flow rather than the more complicated flow regimes expected for $Re \geq 100$ (ref. ¹⁵). The typical procedure described here is for simple trimesic acid (TA), but is similar to that for other systems discussed later (see Fig. 3 and Supplementary Information section 2.2).

In brief, we start by mixing an undersaturated solution of a crystallizing substance in a solvent (47 mg of TA per 0.4 ml of dimethylformamide

(DMF)) with 0.35 ml of the same solvent containing a polyionic liquid polymer (300 mg of poly(3-cyanomethyl-1-vinylimidazolium bis(trifluoromethanesulfonyl)imide); henceforth PIL-1, molecular weight $4.02 \times 10^5 \text{ g mol}^{-1}$, Fig. 2; Supplementary Video 2). The concentration of TA in the 0.75 ml of solution thus prepared exceeds the saturation level by 15.5 mg, even though approximately twice as much TA can be dissolved in pure DMF (92.6 mg per 0.75 ml), which indicates that PIL-1 and TA are competing for shared DMF. The TA/PIL-1/DMF mixture is poured into the Couette cell and, when the inner cylinder begins to rotate, is subject to uniform shear flow (Supplementary Information section 4.4). First, needle-shaped crystals become visible to the naked eye after about 30 s of rotation, move with the fluid, and gradually grow to about 440 μm after 10 min and about 740 μm after one hour (Fig. 1c and blue line in Fig. 1d). Spectroscopic signatures (from powder X-ray diffraction and ^1H nuclear magnetic resonance (NMR) of washed and

¹Center for Soft and Living Matter, Institute for Basic Science, Ulsan, South Korea. ²Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan, South Korea. ³School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing, China. ⁴Department of Materials and Environmental Chemistry, Stockholm University, Stockholm, Sweden.

⁵Department of Applied Chemistry, School of Natural and Applied Sciences, Northwestern Polytechnical University, Xi'an, China. ⁶These authors contributed equally: Jian-Ke Sun, Yaroslav I. Sobolev. ✉e-mail: nanogryzbowski@gmail.com

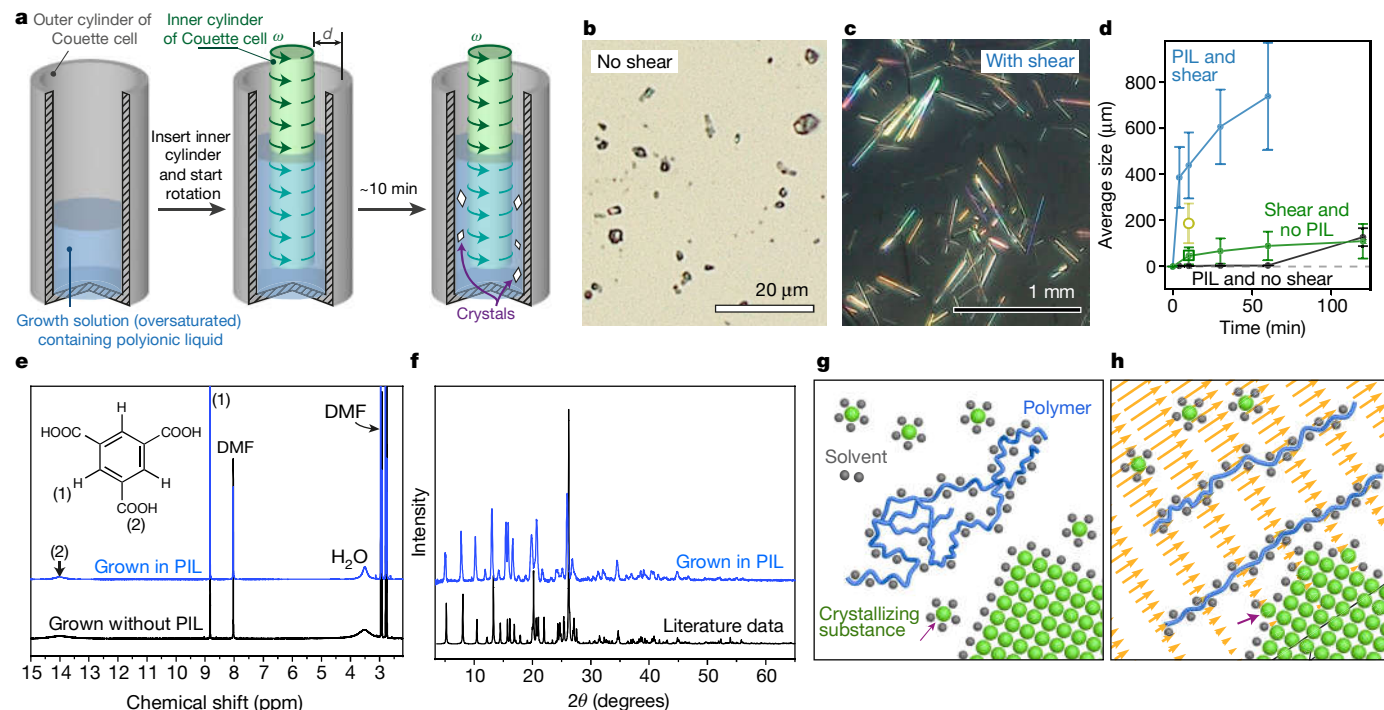


Fig. 1 | Shear-enhanced growth of TA crystals in the presence of an ionic polymer. **a**, Illustration of experimental set-up. **b**, **c**, Optical images of TA crystals grown within 10 min from the same TA/PIL-1/DMF solution without (**b**) and with (**c**) applied shear. We note that the scale bars are different. Image **b** is taken with bright-field illumination; image **c** is taken with crossed polarizers. **d**, Average sizes of TA crystals increasing with time under shear in the presence of PIL-1 (blue; statistics based on $n = 49$ –58 crystals analysed for each time point), without shear but in the presence of PIL-1 (black; $n = 35$ –43), and with shear in pure DMF (green; $n = 236$ –391). Error bars indicate standard deviations of sizes, not the errors of the mean. The green line corresponds to control experiments without PIL-1 but under shear, in which 15.5 mg oversaturation of TA in DMF was achieved by adding 108.1 mg of TA powder to 0.75 ml of pure DMF before application of shear (we verified separately that the saturation level of TA in pure DMF is 92.6 mg per 0.75 ml). The outcome was not sensitive to the time it took most of the TA powder to dissolve: the average size of the crystals obtained was the same in another experiment whereby 85 mg of TA was first completely dissolved in 0.75 ml of pure DMF at 25 °C, and an additional 23 mg of TA powder was added to the solution immediately before the start of cell rotation (open green square). Finally, the open yellow circle corresponds to

an experiment in which 108.1 mg of TA (but no PIL-1) was first completely dissolved in 0.75 ml of DMF at a slightly elevated temperature of 36 °C and then subjected to shear at 25 °C for 10 min. In other words, oversaturation in this experiment was achieved by cooling. In all experiments with non-zero shear, the mean shear rate was $\dot{\gamma} = 167 \text{ s}^{-1}$. **e**, ^1H NMR spectra of the washed and redissolved TA single crystals grown in PIL-1/DMF at $\dot{\gamma} = 167 \text{ s}^{-1}$ shear flow (blue) and by evaporation of DMF from the TA/DMF solution without any PIL (black). The chemical structure in the inset is TA. **f**, Powder X-ray diffraction spectra of TA crystals grown in PIL-1/DMF at $\dot{\gamma} = 167 \text{ s}^{-1}$ shear flow (blue) compared to literature data⁴⁰ for pure TA crystals grown in DMF (black). **g**, **h**, Scheme of the conjectured mechanism: the polymer (blue) and the crystallizing substance (green) compete for shared solvent (grey). When shear flow is absent (**g**), the amount of solvent sufficient to solvate the entangled polymer is lower than that to solvate the polymer disentangled by shear flow (**h**). The disentangling polymer ‘steals’ this additional solvent from the molecules of the crystallizing substance, causing these molecules to attach to the nearby crystal (the purple arrow in both panels points to the same particle). Orange arrows illustrate the vector field of fluid velocity in the case of a mean shear flow. For effects of shear on nucleation, see the discussion in the main text and Fig. 4.

redissolved crystals in Fig. 1e, f; see also Fourier-transform infrared spectroscopy in Supplementary Fig. 13 and single-crystal X-ray diffraction in Supplementary Information section 2), are free of any PIL-1, match the spectra of TA-DMF crystals reported in the literature, and are of crystallographic quality (in terms of single-crystal X-ray diffraction) as good as those of TA crystals grown by conventional recrystallization or solvent evaporation (Supplementary Fig. 12c). In sharp contrast, if the same solution or protocol is used but no rotation is applied, the ill-shaped (Fig. 1b) crystals are only about 2 μm long after 10 min of growth (black line in Fig. 1d). When PIL-1 is absent but shear is applied, the needle-like crystals are about 44 μm long at 10 min (green line in Fig. 1d). None of these and also none of some other control experiments summarized in Fig. 1d (see also Fig. 4 for growth from powders using different shear rates, monomers or polymers of different length) yield crystals of sizes comparable to those grown under shear and with PIL-1 present.

Importantly, similar growth enhancement is observed for other, structurally diverse substances. This claim is supported by the size comparisons (shear + PIL versus no-shear + PIL, same growth times, all at room temperature) in Fig. 3a as well as the corresponding images

of crystals of various small molecules, inorganic salts, metal–organic complexes, and even some proteins (Fig. 3b and Supplementary Figs. 12–35). The average increase in the longest linear dimensions of the crystals is about 16-fold, as high as 42-fold for NaI, 171-fold for TA, and never smaller than 2-fold. The phases and crystallinities of the crystals match those grown over much longer times via traditional solvent evaporation (crystallinities are annotated as percentages in Fig. 3a and plotted in Supplementary Fig. 12; powder X-ray diffraction spectra are shown in Supplementary Figs. 16–28 and 31–35). BET surface areas of porous functional materials are improved with respect to synthesis by conventional methods (without PIL) by 51% for both the porous organic cage¹⁷ and the covalent organic framework²⁰, and by 24% for the metal–organic framework¹⁹. Although absolute values generally depend on the synthetic protocol and activation method, the increase we observe is systematic (with the same activation method for each pair; see Supplementary Fig. 41), and may reasonably be attributed to defects present in the samples grown under shear¹⁶.

Regarding the choice of polymers used in the growth experiments (Fig. 2a), uncharged ones—such as poly(methyl metacrylate), PMMA, or polyvinylidene fluoride, PVDF—also give similar results

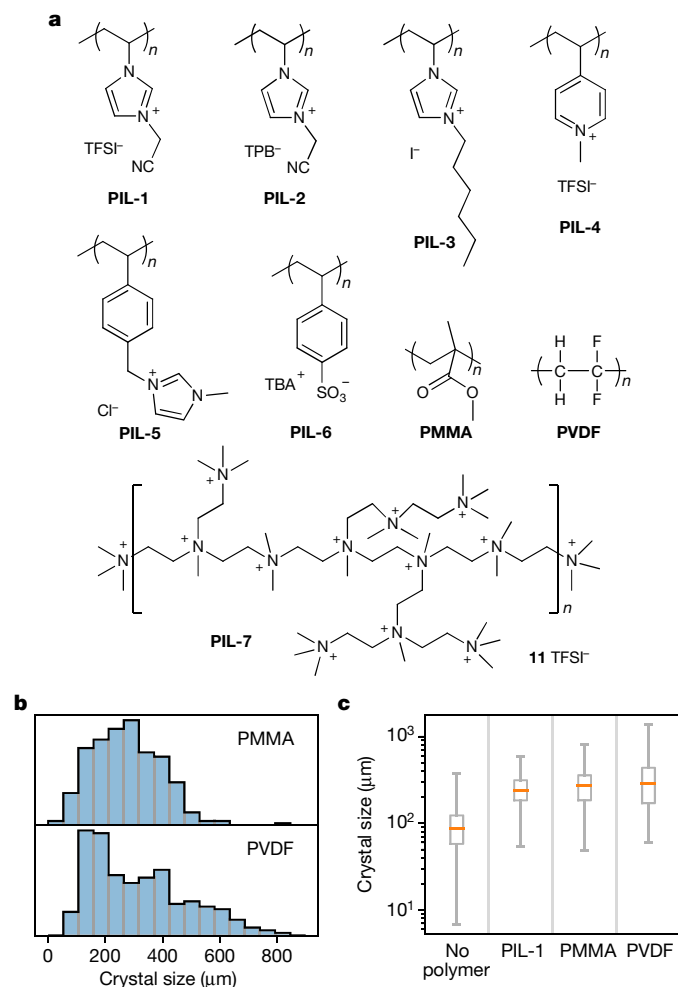


Fig. 2 | Various polymers used for shear-enhanced crystallization. **a**, Most reliable results were obtained using polyionic liquid polymers bearing either positive (PIL-1–PIL-5 and PIL-7) or negative (PIL-6) charges. PIL-1 is poly(3-cyanomethyl-1-vinylimidazolium bis(trifluoromethanesulfonyl)imide); PIL-2 is poly(3-cyanomethyl-1-vinylimidazolium tetrphenylborate); PIL-3 is poly(3-hexyl-1-vinylimidazolium iodide); PIL-4 is poly(1-methyl-4-vinylpyridinium bis(trifluoromethanesulfonyl)imide); PIL-5 is poly(3-methyl-1-(4-vinylbenzyl)imidazolium chloride); PIL-6 is poly(tetrabutylammonium 4-styrenesulfonate); and PIL-7 is quaternary ammonium polyethyleneimine. For synthetic details and characterization, see Supplementary Information section 1. PMMA and PVDF were also used but were not suitable for all solutes, with some of which (for example, TA and NaI) they gelled. (11 refers to NaI.) **b**, Representative size distributions of TA crystals grown from approximately 2- μ m TA powder in PMMA/DMF and PVDF/DMF. **c**, Box plots of crystal sizes for the systems in **b** as well as for the PIL-1 and no-polymer conditions. For all cases, the concentration of polymer was 75 mg per 0.75 ml, growth time 3 h, and mean shear rate $\dot{\gamma} = 85 \text{ s}^{-1}$ (that is, twice as low as in Fig. 1; for effects of $\dot{\gamma}$ on crystal size, see Fig. 4). Molecular weights were 996 kg mol^{-1} (PMMA), 275 kg mol^{-1} (PVDF), and 402 kg mol^{-1} (PIL-1). In **c**, the elements of the box plots are: 25% and 75% quartiles (edges of the boxes), median (midlines), and maximum/minimum values (whiskers). Numbers of crystals on which these statistics are based are: no PIL, 248; PIL-1, 235; PMMA, 284; PVDF, 242.

(see Fig. 2b, c), although with some solutes (for example, TA and NaI) they gelate. Both negatively and positively charged PILs are more robust, confirming previous reports that ionic liquids are versatile solvents compatible with a wide range of solutes^{17,18} and suitable for growing crystals¹⁹ (though never before in shear flow). Conveniently, with the selection of PILs shown in Fig. 2a, we have been able to make our method compatible with solvents ranging from polar (DMF, dimethyl

sulfoxide (DMSO), water and methanol) to less polar (dichloromethane, DCM) by changing the polarity (or sign) of the pendant-chain charges or by varying counterions from small halogen anions to the large tetrabutylammonium (TBA) cation (see specific experimental procedures in Supplementary Information section 2.2).

To better understand the mechanism of shear-enhanced crystal growth, we performed a series of experiments in which we systematically varied the shear rates and polymer chain length. In these experiments, summarized in Fig. 4, we aimed to eliminate any effects of initial nucleation, whose dependence on shear may be convoluted and whose mechanisms are still unclear (see refs.^{10,11} and references therein). Accordingly, we grew the crystals by ripening of TA powders (average particle size $2 \pm 0.5 \mu\text{m}$) in PIL-1/DMF, in contrast to growing the crystals from uniform solutions (see above and Figs. 1, 3). The histogram in Fig. 4a provides evidence that, when other parameters are kept constant, the size of the crystals increases with increasing shear rate and already at $\dot{\gamma} = 167 \text{ s}^{-1}$ further growth becomes limited by the 1-mm gap between the walls of the Couette cell. The distributions in Fig. 4b demonstrate that at a given shear rate, for the same concentration of PIL-1 monomers and same excess (15.5 mg per 0.75 ml of DMF) of TA mass over saturation level (Fig. 4c), crystal sizes increase when the lengths of the PIL-1 polymer chains increase. In parallel, the rheological data in Fig. 4d demonstrate that the viscosities of the solutions also increase in the same order, while the solubility c_s of TA plotted in Fig. 4c decreases. This is an important observation as it strongly indicates that the phenomena we describe cannot be rationalized simply by mixing accelerating transport between crystallites and thus facilitating Ostwald ripening. Both diffusion of macromolecules and mixing slow down in viscous media, whereas the diffusion of small molecules in polymer solutions is either unaffected by increased polymer length or slows down slightly^{20,21}. The solubility c_s of TA in PIL-1/DMF mixtures is at least $(3.5)^{3/7} \approx 1.7$ times slower than in pure DMF and $(1.7)^{3/7} \approx 1.25$ times slower than in monomer/DMF solution (opposite to the actual trend in Fig. 4b) and should slow down with (or, at least, not be affected by) the polymer's molecular weight, which is not the case. We also ruled out potential thermal effects from viscous heating (see Supplementary Information section 4.3). Finally, the curves in Fig. 4d, e capture the decrease of viscosity with increasing shear rate. This so-called 'shear thinning' can be ascribed to polymers or their aggregates disentangling or unfolding in elongational and shear flows; this phenomenon was predicted by de Gennes²² and experimentally confirmed by analyses of bulk properties^{23–26} and even by direct single-molecule imaging²⁷. We note that the decrease in viscosity shown in Fig. 4d, e, owing to some changes of polymer solution microstructure, necessarily begins at shear rates below $10 \div 100 \text{ s}^{-1}$, and therefore these local microstructure changes must be activated by the shear ($25 \div 167 \text{ s}^{-1}$) that we apply in crystal growth experiments (for a detailed discussion of rheology, see Supplementary Information section 6). In this context, we observe that prolonged shearing of the PIL-1/DMF solution reduces the number-averaged hydrodynamic radius r_h observed by DLS (for example, from about 800 nm to 60 nm at $\dot{\gamma} = 167 \text{ s}^{-1}$ for 3 h), although these experiments themselves do not provide detailed insight into the microscopic changes in the polymer's structure (disentanglement, unfolding, breaking of gel-like structures or destruction of the aggregates the PILs are known to form²⁸ and which may break under shear flow²⁹).

All of these observations substantiate a plausible mechanism involving (1) shear-dependent solvation changes and (2) differences in the local shear rates around larger (rather than smaller) crystals. Specifically, as the polymer chains disentangle or their aggregates distort or break under shear, they become better exposed to the solvent and

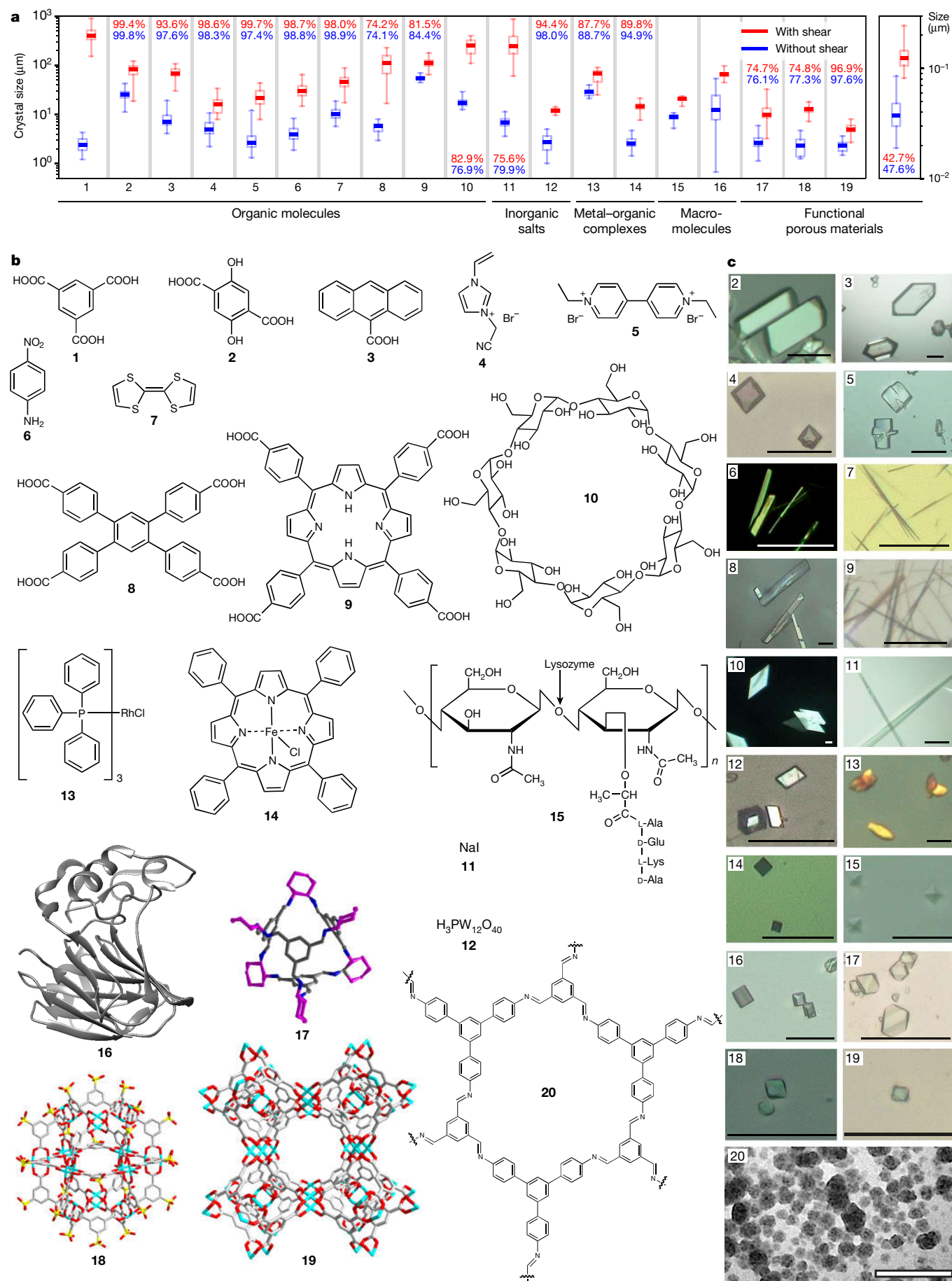


Fig. 3 | See next page for caption.

Fig. 3 | Shear-enhanced growth of additional 19 different crystals in the presence of polyionic liquids. **a**, Sizes of crystals (note logarithmic vertical scale) obtained with shear (red) and without shear (blue) under otherwise identical conditions. The elements of the box plots are: 25% and 75% quartiles (edges of the box), median (midline), and maximum/minimum values (whiskers). Numbers of crystals analysed for each substance are given below (along with experimental conditions) and denoted as n_1 for with-stirring conditions and n_2 for no-stirring conditions. Crystal quality (percentages of crystalline phase evaluated by powder X-ray diffraction analysis) is indicated by numbers on bar plots, in red font for crystals grown with PIL under shear and in blue font for crystals grown with a conventional method (see Supplementary Figs. 11, 12 for details). **b, c**, The specific substances we tested (**b**) and the crystals they typically grow under shear (scale bars = 50 μm , with the exception of **20**, where the scale bar is 0.5 μm) (**c**). All crystals were grown in a Couette cell (Fig. 1a; see Supplementary Information section 5 for cell design) with gap $d = 1 \text{ mm}$ and at a constant shear rate $\dot{\gamma} = 167 \text{ s}^{-1}$ ($\omega = 400 \text{ rpm}$). Various PILs (both positively and negatively charged) were used (chosen to be miscible with the solute or solvent) and their molar concentrations are given in terms of repeat units. Unless otherwise specified, the solvent was DMF. C values are the concentrations of the crystallizing solutes. Times are those of growth under shear and were chosen such that the formation of first crystals could be discerned, in many cases by the naked eye. Detailed experimental conditions, size distributions and additional images of crystals grown with and without shear can be found in Supplementary Information section 2. **1**, TA, growing

time 10 min ($C = 0.30 \text{ M}$, $C_{\text{PIL-1}} = 1 \text{ mM}$, $n_1 = 58$, $n_2 = 35$). **2**, 2,5-dihydroxyterephthalic acid, 10 min ($C = 0.2 \text{ M}$, $C_{\text{PIL-1}} = 0.97 \text{ M}$; $n_1 = 47$, $n_2 = 44$). **3**, Anthracene-9-carboxylic acid, 20 min ($C = 0.45 \text{ M}$, $C_{\text{PIL-1}} = 1.45 \text{ M}$; $n_1 = 30$, $n_2 = 37$). **4**, 3-cyanomethyl-1-vinylimidazolium bromide, 5 min ($C = 0.47 \text{ M}$, $C_{\text{PIL-6}} = 1.5 \text{ M}$, methanol solvent; $n_1 = 25$, $n_2 = 25$). **5**, Ethyl viologen dibromide, 10 min ($C = 0.24 \text{ M}$, $C_{\text{PIL-6}} = 1.17 \text{ M}$, methanol; $n_1 = 43$, $n_2 = 77$). **6**, *p*-Nitroaniline, 10 h ($C = 2.90 \text{ M}$, $C_{\text{PIL-1}} = 2.42 \text{ M}$; $n_1 = 73$, $n_2 = 33$). **7**, Tetrathiafulvalene, 5 min ($C = 0.08 \text{ M}$, $C_{\text{PIL-3}} = 0.01 \text{ M}$, 1:1 v/v DCM/methanol; $n_1 = 71$, $n_2 = 90$). **8**, 1,2,4,5-Tetrakis(4-carboxyphenyl)benzene, 2 h ($C = 0.05 \text{ M}$, $C_{\text{PIL-7}} = 0.04 \text{ M}$; $n_1 = 58$, $n_2 = 36$). **9**, *meso*-tetra(carboxyphenyl)porphyrin, 10 min ($C = 0.005 \text{ M}$, $C_{\text{PIL-1}} = 0.38 \text{ M}$; $n_1 = 35$, $n_2 = 25$). **10**, β -Cyclodextrin, 10 h ($C = 0.09 \text{ M}$, $C_{\text{PIL-1}} = 0.91 \text{ M}$; $n_1 = 26$, $n_2 = 13$). **11**, Sodium iodide, 20 min ($C = 0.47 \text{ M}$, $C_{\text{PIL-3}} = 1.51 \text{ M}$; $n_1 = 49$, $n_2 = 57$). **12**, Phosphotungstic acid, 3 h ($C = 0.06 \text{ M}$, $C_{\text{PIL-5}} = 0.71 \text{ M}$, water; $n_1 = 11$, $n_2 = 22$). **13**, Rhodium(I) tris(triphenylphosphine) chloride, 6 min ($C = 0.04 \text{ M}$, $C_{\text{PIL-3}} = 0.49 \text{ M}$, DCM; $n_1 = 35$, $n_2 = 43$). **14**, Iron(III) *meso*-tetraphenylporphyrin chloride, 3 min ($C = 0.02 \text{ M}$, $C_{\text{PIL-6}} = 1.17 \text{ M}$, DCM; $n_1 = 43$, $n_2 = 29$). **15**, Hen egg white lysozyme, 2 h ($C = 0.002 \text{ M}$, $C_{\text{PIL-5}} = 1.83 \text{ M}$, NaAc-HAc buffer; $n_1 = 21$, $n_2 = 31$). **16**, Thaumatin, 16 h ($C = 0.005 \text{ M}$, $C_{\text{PIL-5}} = 0.71 \text{ M}$, ADA buffer; $n_1 = 99$, $n_2 = 54$). **17**, CC3-R porous molecular cage, 10 min ($C = 0.01 \text{ M}$, $C_{\text{PIL-3}} = 0.54 \text{ M}$, DCM; $n_1 = 57$, $n_2 = 47$). **18**, metal-organic polyhedron (MOP), 10 min ($C = 0.003 \text{ M}$, $C_{\text{PIL-5}} = 1.42 \text{ M}$, methanol; $n_1 = 40$, $n_2 = 14$). **19**, HKUST-1 MOF grown from TA and $\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$, 10 h ($C_{\text{TA}} = 0.27 \text{ M}$, $C_{\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}} = 0.33 \text{ M}$, $C_{\text{PIL-4}} = 0.33 \text{ M}$; $n_1 = 32$, $n_2 = 31$). **20**, TAPB-BTCA covalent organic framework, 15 min ($C_{\text{TAPB}} = 0.02 \text{ M}$, $C_{\text{BTCA}} = 0.02 \text{ M}$, $C_{\text{PIL-1}} = 0.12 \text{ M}$, DMSO; $n_1 = 112$, $n_2 = 112$).

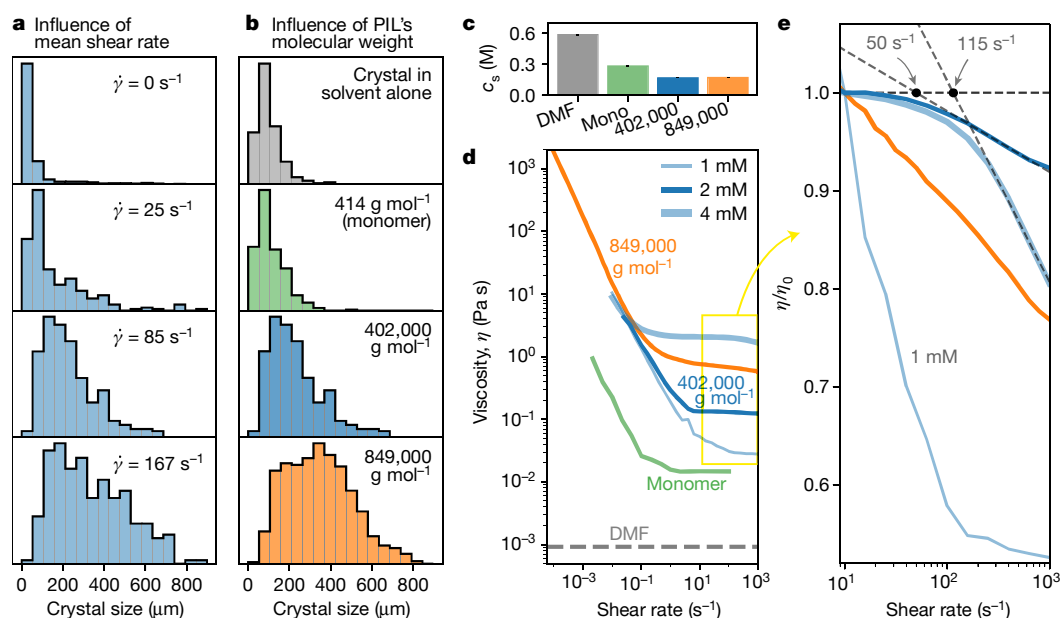


Fig. 4 | Effects of shear rates and polymer's chain length on crystal growth. **a**, Distributions of sizes of the TA crystals grown for 3 h from the same TA powder (not to be confused with experiments on growth from the oversaturated solutions presented in Figs. 1, 3) in PIL-1/DMF (PIL-1 molecular weight 402 kg mol^{-1}) under different mean shear rates. Each histogram is based on the analysis of (top to bottom panels) 1,000, 235, 100 and 131 crystals. **b**, Distributions of sizes of TA crystals grown for 3 h from the same TA powder under $\dot{\gamma} = 85 \text{ s}^{-1}$ mean shear rate in pure DMF (top panel), in monomer of PIL-1 in DMF (300 mg per 0.75 ml), and in PIL-1/DMF (300 mg per 0.75 ml) with two different molecular weights (indicated). Each histogram is based on the analysis of, top to bottom panels, 315, 348, 100 and 383 crystals. **c**, Equilibrium solubility (molar concentration at saturation) of TA in the same liquids as in panel **b**, measured with no shear. **d, e**, Measurements of viscosity at different shear rates. There is strong 'primary' shear thinning at low shear rates (10^{-4} s^{-1} to 10 s^{-1} , **d**), and a weaker 'secondary' shear thinning at higher shear rates ($>10 \text{ s}^{-1}$, **e**), both are facilitated by addition of solvent (compare the blue curves measured for the same molecular weight of 402 kg mol^{-1} , but at different concentrations, as indicated in the legend) or by increasing the molecular weight (compare orange curve of PIL-1 with molecular weight 849 kg mol^{-1} at

concentration 300 mg per 0.75 ml DMF (same as in the crystal growth experiments) to dark blue curve for molecular weight 402 kg mol^{-1} at 300 mg per 0.75 ml DMF (2 mM). As more solvent is added, the 'primary' shear thinning (**d**) progresses further into higher shear rates and reaches lower viscosities, indicating that disentanglement is limited by available solvent. In the high shear region (**e**), 'secondary' thinning also becomes more prominent when more solvent is added or when molecular weight is increased, and whenever a clear plateau could be discerned, onset of thinning (shown by black dots, which are intersections of dashed asymptotes) shifts to lower shears with more dilution (50 s^{-1} for 2 mM versus 115 s^{-1} for 1 mM). For a more dilute, 1 mM sample, the 'secondary' thinning (**e**) begins at an even lower shear rate and merges into the first, low-shear thinning process, thus showing no clear plateau of viscosity. Curves from the yellow rectangle in **d** are plotted in **e** on a semi-logarithmic scale with viscosity η for each curve normalized by its value η_0 at 9 s^{-1} . Viscosity of pure DMF (grey dashed line) is taken from literature³⁹. The curve for the monomer (green) is measured at higher concentration (2 g/0.75 ml DMF) than for PILs, since more dilute monomer solution has even lower viscosity and could not be reliably measured by our rheometer.

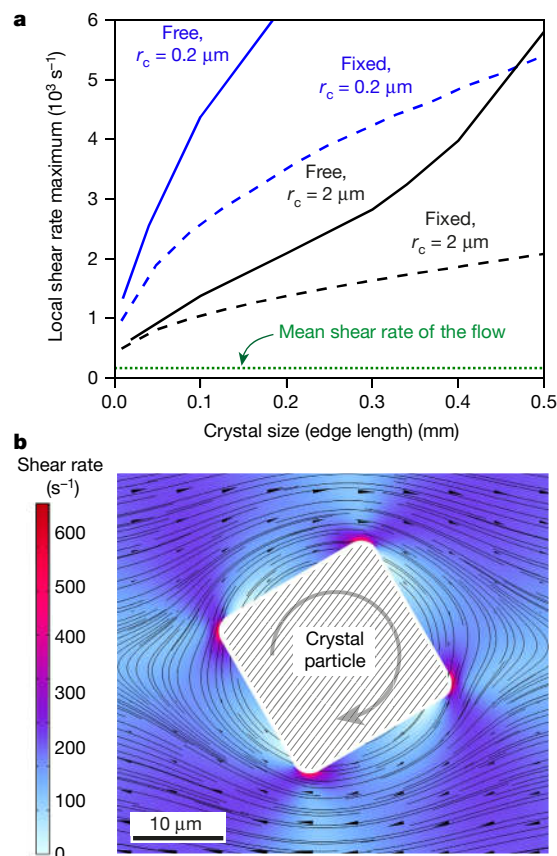


Fig. 5 | Effects of particle size on local shear rates. a, Maximum values of shear rate near an object (a rod as in panel **b**) in Couette flow ($d = 1$ mm, mean $\dot{\gamma} = 167$ s $^{-1}$ indicated by the horizontal dotted green line) as a function of the object's size and corner sharpness (r_c is the radius of corner curvature). Dashed curves were calculated for a non-rotating rod (with two of its faces kept parallel to the cell walls). Solid curves were calculated for a rod that is being freely rotated by the flow. **b**, Theoretical map of shear rate near a long, freely rotating rod in horizontal Couette flow (distance between walls $d = 1$ mm, mean $\dot{\gamma} = 167$ s $^{-1}$). The rod's axis points into the page; the rod's cross-section is a square (20 μ m per edge) with rounded corners ($r_c = 2$ μ m). Being freely rotated by the shear flow, the rod maintains a constant angular velocity (clockwise, as indicated by the circular grey arrow). Black curves are streamlines. Black cones indicate velocity direction and magnitude. This map depicts one instant of time; as the object rotates, maximal local shear rate oscillates in phase with this rotation, but regions of high shear rate remain mostly localized near the sharp corners at all times. The liquid is assumed to be newtonian, although very similar results were obtained when we used realistic dependence of viscosity on shear rate (from Fig. 4d, solid blue curve). Details of all these calculations are described in Supplementary Information section 4 (see also Supplementary

their effective volume of solvation layer^{30,31} increases. This phenomenon is in line with previous experiments on both polymers and ionic liquids²⁹, whereby adding more solvent enhances shear thinning, hinting that shear may cause greater demand for solvent by the polymer or ionic liquid; we found that this trend also holds for PIL-1/DMF (see blue curves in Fig. 4d). What this means for our experiments is that the disentangled polymers can effectively compete for solvent with the solute which, upon losing the solvent, starts to crystallize (this effect is loosely analogous to 'salting out', which causes the precipitation and crystallization^{32,33} of various solutes, including biomolecules^{34,35}). We observe that this mechanism explains why some other polymers are less robust or less general than PILs (such as PMMA and PVDF; see Fig. 2b, c), because they gelate or precipitate before the solute can develop large crystals. By contrast, PILs do not precipitate easily and are known to be versatile solvents, not unlike molten salts^{17,18}.

The second part of the proposed mechanism rationalizes the preferential growth of larger (rather than smaller) crystals which, as we mentioned above in the context of crystal sizes increasing with viscosity and decreasing with solubility, cannot be ascribed to classical Ostwald ripening. Instead, the explanation might lie in the fact that in a mean shear flow, the local shear near a particle with sharp edges increases with particle size^{36,37}. For our system, such an increase is supported by the results of computational fluid dynamics simulations of liquid flows around rod-shaped particles with square cross-sections ranging from 2 μ m to 1 mm and moving freely in a Couette cell (Fig. 5b, c). Therefore, the disentanglement of PILs and 'competition' for solvent are expected to be more pronounced near larger particles, which therefore grow preferentially; conversely, the smaller particles remain more soluble. Experimentally, we found that a mean shear of $\dot{\gamma} = 167$ s $^{-1}$ indeed decreases global solubility of TA in PIL-1/DMF (molecular weight 8.49×10^5 g mol $^{-1}$) from 167.2 ± 2.8 mM to 161.6 ± 0.8 mM (see Supplementary Table 1 in Supplementary Information section 2.1). We note that this 1%–5% modulation of solubility is an order of magnitude larger than the typical values sufficient to drive Ostwald ripening (for example, <0.1% for 10- μ m crystals of TA, assuming a surface energy of 100 mJ m $^{-2}$ in the Ostwald–Freundlich/Kelvin equation).

However, if it were just simple disentanglement or single-chain unfolding, it would have been reversible, and therefore some portion of crystals such as TA would re-dissolve shortly after stopping the rotation of the Couette cell: PIL-1 microstructure would have returned to its original state, releasing the captured solvent. In experiments, TA crystals did not re-dissolve even 15 h after the rotation had ceased, indicating that relaxation of PILs to the initial state is inhibited—perhaps owing to a gel-like macromolecular crowding in the concentrated solutions we use (this would be in line with onset of shear thinning already at very low shear rates <10 $^{-4}$ s $^{-1}$ shown in Fig. 4d, as well as linear response domain at very low strains (around 0.03) shown in the dynamic relaxation data in Supplementary Fig. 50a).

Finally, if the mechanism we propose for shear-enhanced ripening of powders (Fig. 2b, c and Fig. 4) is correct, it should also apply to the growth starting from oversaturated solutions (Fig. 1 and Fig. 3)—although, in the nucleation phase, it would strongly suppress smaller nuclei in favour of the larger ones, and we would expect large crystals to grow from such solutions faster than from powders. This is indeed the case and solution growth is approximately an order of magnitude faster. For instance, TA crystals reach a size of 440 μ m on average after 10 min of growth from solution (Fig. 1d) versus an average size of 336 μ m after only 3 h of growth from powder (Fig. 4a, bottom panel). Furthermore, the impact of this mechanism during nucleation phase must far outweigh the previously reported effects of shear flow on nucleation^{9–11,38}, since those effects predict the negative effect of shear on the final crystal size—that is, the opposite of what we observe (Fig. 4a).

In summary, we showed that good-quality crystals of various kinds can grow larger and more rapidly when subject to shear in the presence of polymers. Since this trend is observed for crystals and polymers of various types, it can reasonably be explained by physical effects rather than the nuances of specific polymer–solute chemical interactions. At the same time, such interactions might have more subtle effects—we have seen, for instance, that linear PIL-1 and branched PIL-7 can give crystals of the same space group but different habits (see Supplementary Fig. 37). Such effects and also more detailed theoretical models certainly merit further study of this interesting non-equilibrium system. From a practical point of view, we anticipate that our technically straightforward, constant-temperature method will be useful as a means of accelerating crystal growth, especially for substances that must be kept within a narrow temperature range (for example, proteins) or cannot be recrystallized (such as metal–organic frameworks and covalent organic frameworks).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2042-1>.

- Shekunov, B. Y. & York, P. Crystallization processes in pharmaceutical technology and drug delivery design. *J. Cryst. Growth* **211**, 122–136 (2000).
- Variankaval, N., Cote, A. S. & Doherty, M. F. From form to function: crystallization of active pharmaceutical ingredients. *AIChE J.* **54**, 1682–1688 (2008).
- Ulrich, J. & Froberg, P. Problems, potentials and future of industrial crystallization. *Front. Chem. Sci. Eng.* **7**, 1–8 (2013).
- Censi, R. & Di Martino, P. Polymorph impact on the bioavailability and stability of poorly soluble drugs. *Molecules* **20**, 18759–18776 (2015).
- Lee, E. H. A practical guide to pharmaceutical polymorph screening & selection. *Asian J. Pharm. Sci.* **9**, 163–175 (2014).
- ICH Q6A Specifications: Test Procedures And Acceptance Criteria For New Drug Substances And New Drug Products: Chemical Substances Report CPMP/ICH/367/96 (European Medicines Agency, 2000); <https://www.ema.europa.eu/en/ich-q6a-specifications-test-procedures-acceptance-criteria-new-drug-substances-new-drug-products>.
- Guidance For Industry. ANDAS: Pharmaceutical Solid Polymorphism Chemistry, Manufacturing, And Controls Information (Food and Drug Administration, 2007); http://academy.gmp-compliance.org/guidemgr/files/POLYMORPHISM_7590FNL.PDF
- Shamlou, P. A. & Titchener-Hooker, N. Turbulent aggregation and breakup of particles in liquids in stirred vessels. In *Processing of Solid-Liquid Suspensions* 1–25 (Butterworth-Heinemann, 1993).
- Wang, J. & Estrin, J. Secondary nucleation of sucrose by fluid shear in aqueous solutions. *Chem. Eng. Commun.* **152/153**, 275–286 (1996).
- Forsyth, C. et al. Influence of controlled fluid shear on nucleation rates in glycine aqueous solutions. *Cryst. Growth Des.* **15**, 94–102 (2015).
- Forsyth, C., Burns, I. S., Mulheran, P. A. & Sefcik, J. Scaling of glycine nucleation kinetics with shear rate and glass–liquid interfacial area. *Cryst. Growth Des.* **16**, 136–144 (2016).
- Botsaris, G. D. Secondary nucleation—a review. In *Industrial Crystallization* 3–22 (Springer, 1976).
- McCabe, W. L., Smith, J. C. & Harriott, P. *Unit Operations Of Chemical Engineering* 1130 (McGraw-Hill, 1993).
- Sung, C. Y., Estrin, J. & Youngquist, G. R. Secondary nucleation of magnesium sulfate by fluid shear. *AIChE J.* **19**, 957–962 (1973).
- Andereck, C. D., Liu, S. S. & Swinney, H. L. Flow regimes in a circular Couette system with independently rotating cylinders. *J. Fluid Mech.* **164**, 155–183 (1986).
- Hasell, T., Chong, S. Y., Jelfs, K. E., Adams, D. J. & Cooper, A. I. Porous organic cage nanocrystals by solution mixing. *J. Am. Chem. Soc.* **134**, 588–598 (2012).
- Antonietti, M., Kuang, D., Smarsly, B. & Zhou, Y. Ionic liquids for the convenient synthesis of functional nanoparticles and other inorganic nanostructures. *Angew. Chem. Int. Ed.* **43**, 4988–4992 (2004).
- Zhen, M., Yu, J. & Dai, S. Preparation of inorganic materials using ionic liquids. *Adv. Mater.* **22**, 261–285 (2010).
- Gao, M. R., Yu, S. H., Yuan, J., Zhang, W. & Antonietti, M. Poly(ionic liquid)-mediated morphogenesis of bismuth sulfide with a tunable band gap and enhanced electrocatalytic properties. *Angew. Chem. Int. Ed.* **55**, 12812–12816 (2016).
- Berry, G. C. & Fox, T. G. The viscosity of polymers and their concentrated solutions. *Fortsch. Hochpolym. Adv. Polymer Sci.* **5**, 261–357 (1968).
- Szymański, J., Patkowski, A., Wilk, A., Garstecki, P. & Holyst, R. Diffusion and viscosity in a crowded environment: from nano-to macroscale. *J. Phys. Chem. B* **110**, 25593–25597 (2006).
- De Gennes, P. G. Coil-stretch transition of dilute flexible polymers under ultrahigh velocity gradients. *J. Chem. Phys.* **60**, 5030–5042 (1974).
- Cottrell, F. R., Merrill, E. W. & Smith, K. A. Conformation of polyisobutylene in dilute solution subjected to a hydrodynamic shear field. *J. Polym. Sci. A* **7**, 1415–1434 (1969).
- Larson, R. G. *Constitutive Equations For Polymer Melts And Solutions* (Butterworth-Heinemann, 1988).
- Link, A. & Springer, J. Light scattering from dilute polymer solutions in shear flow. *Macromolecules* **26**, 464–471 (1993).
- Lee, E. C., Solomon, M. J. & Muller, S. J. Molecular orientation and deformation of polymer solutions under shear: a flow light scattering study. *Macromolecules* **30**, 7313–7321 (1997).
- Smith, D. E., Babcock, H. P. & Chu, S. Single polymer dynamics in steady shear flow. *Science* **283**, 1724–1727 (1999).
- Sun, J. K. et al. General synthetic route toward highly dispersed metal clusters enabled by poly(ionic liquid)s. *J. Am. Chem. Soc.* **139**, 8971–8976 (2017).
- Burrell, G. L., Dunlop, N. F. & Separovic, F. Non-Newtonian viscous shear thinning in ionic liquids. *Soft Matter* **6**, 2080–2086 (2010).
- Del Gaudio, S. & Amadei, A. The unfolding effects on the protein hydration shell and partial molar volume: a computational study. *Phys. Chem. Chem. Phys.* **18**, 28175–28182 (2016).
- Norman, A. I., Yiwei, F., Ho, D. L. & Greer, S. C. Folding and unfolding of polymer helices in solution. *Macromolecules* **40**, 2559–2567 (2007).
- Noyes, A. A. The physical properties of aqueous salt solutions in relation to the ionic theory. *Science* **20**, 577–587 (1904).
- Lewis, G. N. & Randall, M. *Thermodynamics And The Free Energy Of Chemical Substances* (McGraw-Hill, 1923).
- Cohn, E. J. The physical chemistry of the proteins. *Physiol. Rev.* **5**, 349–437 (1925).
- Miller, S. A., Dykes, D. D. & Polesky, H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
- Schmid, D. W. Finite and infinite heterogeneities under pure and simple shear. PhD thesis, ETH Zurich (2002).
- Nesbitt, W. S. et al. A shear gradient-dependent platelet aggregation mechanism drives thrombus formation. *Nat. Med.* **15**, 665–673 (2009).
- Ramel, P. R., Campos, R. & Marangoni, A. G. Effects of shear and cooling rate on the crystallization behavior and structure of cocoa butter: shear applied during the early stages of nucleation. *Cryst. Growth Des.* **18**, 1002–1011 (2018).
- Marcus, Y. *The Properties Of Solvents* (Wiley, 1998).
- Chatterjee, S., Pediredi, V. R., Ranganathan, A. & Rao, C. N. R. Self-assembled four-membered networks of trimesic acid forming organic channel structures. *J. Mol. Struct.* **520**, 107–115 (2000).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Data availability

All data used in the calculations are available in GitHub repository (<https://doi.org/10.5281/zenodo.3533635>).

Code availability

All computer codes and COMSOL project files used in the calculations are available in GitHub repository (<https://doi.org/10.5281/zenodo.3533635>).

Acknowledgements We acknowledge support from the Institute for Basic Science Korea (Project Code IBS-R020-D1). We thank S. Lach for his help in the design and manufacture of the Couette cells, M. Siek for DLS measurements, C. Cahoon and K. K. Zheng

for rheological measurements, J. Yuan for help with PIL synthesis, and W. Adamkiewicz for discussions.

Author contributions J.-K.S. designed and performed most of the experiments. Y.I.S. developed theoretical models and helped with some experiments. W.Z. and Q.Z. synthesized most of the PIL polymers. B.A.G. conceived and supervised the research. All authors wrote the paper.

Competing interests A patent application based on these results has been submitted by the Institute for Basic Science (South Korea Patent Application 10-2019-0008413; inventors J.-K.S., Y.I.S. and B.A.G.).

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2042-1>.

Correspondence and requests for materials should be addressed to B.A.G.

Peer review information *Nature* thanks Andrew Cooper, Laurence Noirez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Asynchronous carbon sink saturation in African and Amazonian tropical forests

<https://doi.org/10.1038/s41586-020-2035-0>

Received: 9 June 2019

Accepted: 19 December 2019

Published online: 4 March 2020

 Check for updates

A list of authors and affiliations appears at the end of the paper.

Structurally intact tropical forests sequestered about half of the global terrestrial carbon uptake over the 1990s and early 2000s, removing about 15 per cent of anthropogenic carbon dioxide emissions^{1–3}. Climate-driven vegetation models typically predict that this tropical forest ‘carbon sink’ will continue for decades^{4,5}. Here we assess trends in the carbon sink using 244 structurally intact African tropical forests spanning 11 countries, compare them with 321 published plots from Amazonia and investigate the underlying drivers of the trends. The carbon sink in live aboveground biomass in intact African tropical forests has been stable for the three decades to 2015, at 0.66 tonnes of carbon per hectare per year (95 per cent confidence interval 0.53–0.79), in contrast to the long-term decline in Amazonian forests⁶. Therefore the carbon sink responses of Earth’s two largest expanses of tropical forest have diverged. The difference is largely driven by carbon losses from tree mortality, with no detectable multi-decadal trend in Africa and a long-term increase in Amazonia. Both continents show increasing tree growth, consistent with the expected net effect of rising atmospheric carbon dioxide and air temperature^{7–9}. Despite the past stability of the African carbon sink, our most intensively monitored plots suggest a post-2010 increase in carbon losses, delayed compared to Amazonia, indicating asynchronous carbon sink saturation on the two continents. A statistical model including carbon dioxide, temperature, drought and forest dynamics accounts for the observed trends and indicates a long-term future decline in the African sink, whereas the Amazonian sink continues to weaken rapidly. Overall, the uptake of carbon into Earth’s intact tropical forests peaked in the 1990s. Given that the global terrestrial carbon sink is increasing in size, independent observations indicating greater recent carbon uptake into the Northern Hemisphere landmass¹⁰ reinforce our conclusion that the intact tropical forest carbon sink has already peaked. This saturation and ongoing decline of the tropical forest carbon sink has consequences for policies intended to stabilize Earth’s climate.

Tropical forests account for approximately one-third of Earth’s terrestrial gross primary productivity and one-half of Earth’s carbon stored in terrestrial vegetation¹¹. Thus, small biome-wide changes in tree growth and mortality can have global impacts, either buffering or exacerbating the increase in atmospheric CO₂. Models^{2,4,5,7,12}, ground-based observations^{13–15}, airborne atmospheric CO₂ measurements^{3,16}, inferences from remotely sensed data¹⁷ and synthetic approaches^{3,8,18} each suggest that, after accounting for land-use change, the remaining structurally intact tropical forests (that is, those not affected by direct anthropogenic impacts such as logging) are increasing in carbon stocks. This structurally intact tropical forest carbon sink is estimated at approximately 1.2 Pg C yr^{−1} over 1990–2007 using scaled inventory plot measurements¹. Yet, despite its relevance to policy, changes in this key carbon sink remain highly uncertain^{19,20}.

Globally, the terrestrial carbon sink is increasing^{2,7,8,21}. Between 1990 and 2017 the land surface sequestered about 30% of all anthropogenic carbon dioxide emissions^{1,21}. Rising CO₂ concentrations are thought to have boosted photosynthesis more than rising air temperatures have enhanced respiration, resulting in an increasing global terrestrial carbon sink^{2,4,7,8,21}. Yet, for Amazonia, recent results from repeated

censuses of intact forest inventory plots show a progressive two-decade decline in sink strength primarily due to an increase in carbon losses from tree mortality⁶. It is unclear if this simply reflects region-specific drought impacts^{22,23}, or potentially chronic pan-tropical impacts of either heat-related tree mortality^{24,25}, or results from internal forest dynamics as past increases in carbon gains leave the system²⁶. A more recent deceleration of the rate of increase in carbon gains from tree growth is also contributing to the declining Amazon sink⁶. Again, it is not known whether this is a result of either pan-tropical saturation of CO₂ fertilization, or rising air temperatures, or is simply a regional drought impact. To address these uncertainties, we (1) analyse an unprecedented long-term inventory dataset from Africa, (2) pool the new African and existing Amazonian records⁶ to investigate the putative environmental drivers of changes in the tropical forest carbon sink, and (3) project its likely future evolution.

We collected, compiled and analysed data from structurally intact old-growth forests from the African Tropical Rainforest Observation Network²⁷ (217 plots) and other sources (27 plots) spanning the period 1 January 1968 to 31 December 2014 (Extended Data Fig. 1; Supplementary Table 1). In each plot (mean size, 1.1 ha), all trees ≥100 mm in stem

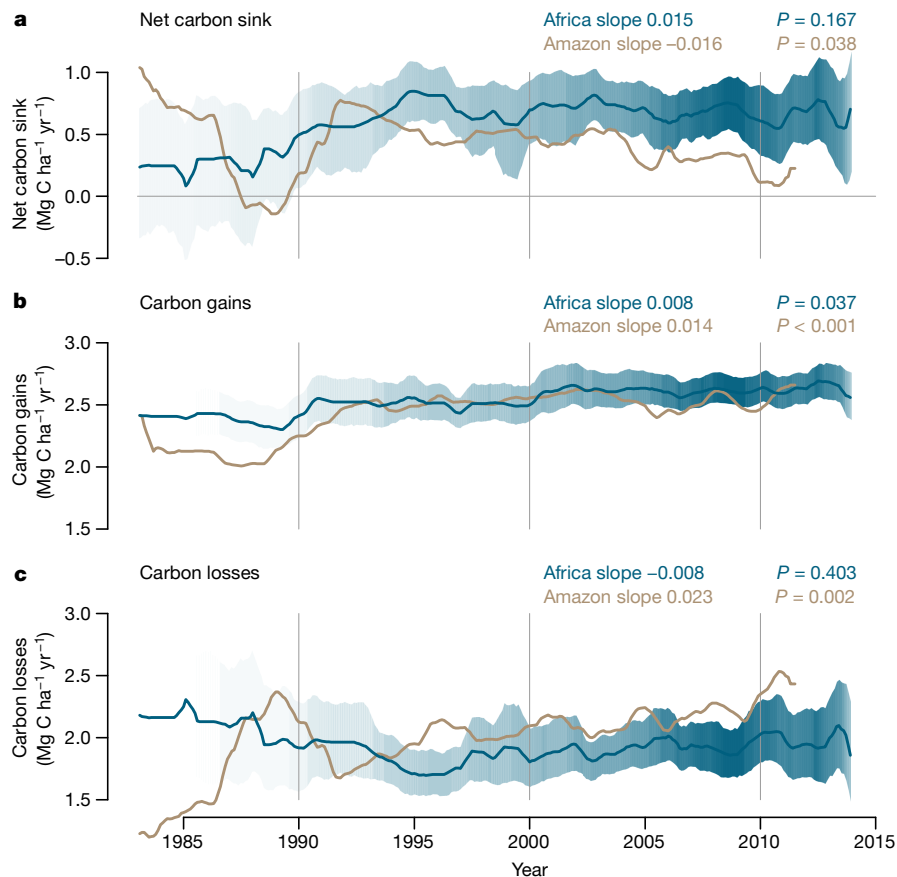


Fig. 1 | Long-term carbon dynamics of structurally intact old-growth tropical forests in Africa and Amazonia. **a–c**, Trends in net aboveground live biomass carbon (**a**), carbon gains to the system from wood production (**b**), and carbon losses from the system from tree mortality (**c**), measured in 244 African inventory plots (blue lines) and contrasting published⁶ Amazonian inventory data (brown lines; 321 plots). For Africa we

show complete years with at least 25 plots monitored; for Amazonia we show the published record⁶. Shading corresponds to the 95% CI, with darker shading indicating a greater number of plots monitored in that year (the lightest shading indicates the minimum 25 plots monitored). The CI for the Amazonian dataset is omitted for clarity, but can be seen in Fig. 3. Slopes and *P* values are from linear mixed effects models (see Methods).

diameter were identified, mapped and measured at least twice using standardized methods (135,625 trees monitored). Live biomass carbon stocks were estimated for each census date, with carbon gains and losses calculated for each interval (Extended Data Fig. 2).

Continental carbon sink trends

We detect no long-term trend in the per unit area African tropical forest carbon sink over three decades to 2015 ($P = 0.167$; Fig. 1). The aboveground live biomass sink averaged 0.66 tonnes of carbon per hectare per year ($0.66 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ with 95% confidence interval (CI) of 0.53–0.79 and $n = 244$) and was significantly greater than zero for every year since 1990 (Fig. 1; $P < 0.001$ for each time period in Table 1). Although very similar to past reports ($0.63 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$)¹³, this first estimate of the temporal trend in Africa contrasts with the significantly declining ($P = 0.038$) Amazonian trend⁶ (Fig. 1). A linear mixed effects model shows a significant difference in the slopes of the sink trends for the two continents over the common time window (pooled data from both continents, common time window, 1 January 1983 to mid-2011; $P = 0.017$). Therefore, the per unit area sink strength of the two largest expanses of tropical forest on Earth diverged in the 1990s and 2000s.

The proximal cause of the divergent sink patterns is a significant increase ($P = 0.002$) in carbon losses (from tree mortality, that is, the loss of carbon from the live biomass pool) in Amazonian forests, with no detectable trend over three decades in African forests ($P = 0.403$; Fig. 1; Table 1). A linear mixed effects model using pooled data shows

a significant difference in slopes of carbon losses between the two continents over the common time window ($P = 0.027$; 1 January 1983 to mid-2011). Long-term trends in carbon gains (from tree growth and newly recruited trees) show significant increases on both continents ($P = 0.037$ for Africa; $P < 0.001$ for Amazon; Fig. 1), and we could detect no difference in slopes between the continents ($P = 0.348$; carbon gains from tree growth alone also show no continental difference in long-term trends, $P = 0.322$). However, an assessment of how underlying environmental drivers affect carbon gains and losses is needed to understand the ultimate causes of the divergent sink patterns.

Understanding the carbon sink trends

We first investigate those environmental drivers exhibiting long-term change that affect photosynthesis and respiration in theory-driven models: atmospheric CO_2 concentration, surface air temperature and water availability. Bivariate models (Fig. 2) and a linear mixed effects model of carbon gains (Extended Data Table 1), with censuses nested within plots, and pooling the new African and published Amazonian data, show a significant positive relationship with CO_2 ($P = 0.021$ in Fig. 2; $P = 0.001$ in Extended Data Table 1), and significant negative relationships with mean annual temperature (MAT; $P < 0.001$ in Fig. 2 and Extended Data Table 1) and drought ($P = 0.003$ in Fig. 2; $P < 0.001$ in Extended Data Table 1), with drought measured as the maximum climatological water deficit (MCWD)¹⁴. These results are consistent with a positive CO_2 fertilization effect, and negative effects

Table 1 | Carbon sink in structurally intact old-growth tropical forests in Africa, Amazonia and the pan-tropics, 1980–2040

Period	Number of plots		Per unit area aboveground live biomass C sink (Mg C ha ⁻¹ yr ⁻¹)			Total C sink (Pg C yr ⁻¹) ^a		
	Africa	Amazon	Africa	Amazon	Pan-tropics ^b	Africa	Amazon	Pan-tropics ^b
1980–1990	45	73	0.33 (0.06–0.63)	0.35 (0.06–0.59)	0.35 (0.07–0.62)	0.28 (0.05–0.53)	0.49 (0.08–0.82)	0.87 (0.16–1.52)
1990–2000	96	172	0.67 (0.43–0.89)	0.53 (0.42–0.65)	0.57 (0.39–0.74)	0.50 (0.32–0.66)	0.68 (0.54–0.83)	1.26 (0.88–1.63)
2000–2010	194	291	0.70 (0.55–0.84)	0.38 (0.26–0.48)	0.50 (0.35–0.64)	0.46 (0.37–0.56)	0.45 (0.31–0.57)	0.99 (0.70–1.25)
2010–2015 ^c	184	172	0.66 (0.40–0.91)	0.24 (0.00–0.47)	0.40 (0.15–0.65)	0.40 (0.24–0.56)	0.27 (0.00–0.52)	0.73 (0.25–1.18)
2010–2020 ^d	–	–	0.63 (0.36–0.89)	0.23 (–0.05–0.50)	0.38 (0.11–0.65)	0.37 (0.21–0.53)	0.25 (–0.05–0.54)	0.68 (0.17–1.16)
2020–2030 ^d	–	–	0.59 (0.24–0.93)	0.12 (–0.29–0.51)	0.30 (–0.08–0.67)	0.31 (0.13–0.49)	0.12 (–0.29–0.52)	0.47 (–0.15–1.07)
2030–2040 ^d	–	–	0.55 (0.08–0.99)	0.00 (–0.54–0.49)	0.21 (–0.29–0.67)	0.26 (0.04–0.47)	0.00 (–0.50–0.46)	0.29 (–0.46–0.97)

This table covers 1 January 1980 to 31 December 2014 and predictions to 31 December 2039. Mean values are in boldface, future predictions in italics, uncertainties in parentheses: 95% bootstrapped confidence intervals for 1980–2015, and 2σ for the predictions (2010–2040).

^aThe total continental C sink is the per unit area aboveground C sink multiplied by intact forest area (from ref. ¹; see Extended Data Table 2) and includes continent-specific estimates of three carbon-stock components that were not measured in the inventory plots: trees with a diameter at breast height of <100 mm, lianas and roots (see Methods).

^bThe per unit area pan-tropical aboveground live biomass C sink is the area-weighted mean of African, Amazonian and Southeast Asian sink values. Southeast Asian values were from published per unit area carbon sink data¹⁵ ($n = 49$ plots) for 1990–2015, with 1980–1990 assumed to be the same as 1990–2000 owing to very low sample sizes. The pan-tropical total C sink is the sum of African, Amazonian and Southeast Asian total continental carbon sink values. The continental sink in Southeast Asia is a modest and declining contribution to the pan-tropical sink, owing to the very small area of intact forest remaining, at 0.11 Pg C yr⁻¹, 0.08 Pg C yr⁻¹, 0.07 Pg C yr⁻¹ and 0.06 Pg C yr⁻¹ in the 1980s, 1990s, 2000s and 2010s, respectively; hence uncertainty in the Southeast Asian sink cannot reverse the pan-tropical declining sink trend.

^cThe Amazonian sink in the 2010–2015 time window was calculated from 172 plots that were mostly measured between 1 January 2010 and mid-2011. The lack of temporal coverage later in this period probably has little impact on the results; adding modelled results for 1 January 2012 to 31 December 2014 gives a per unit area aboveground sink of 0.25 Mg C ha⁻¹ yr⁻¹ (0.00–0.49), which would increase the pan-tropical total C sink by 0.01 Pg C yr⁻¹.

^dPer unit area total C sink for 2010–2020, 2020–2030 and 2030–2040 was predicted using parameters from Table 2, except for the 2010–2020 sink in Africa, which is the mean of the measured sink from 2010–2015 and the modelled sink from 2015–2020. For the Asian sink we assumed the same parameters as for Africa, because Asian forest median CRT is 61 years, close to the African median of 63 years.

of higher temperatures and drought on tree growth, consistent with temperature-dependent increases in autotrophic respiration, and temperature- and drought-dependent reductions in carbon assimilation. By contrast, the equivalent models for carbon losses show no significant relationships with CO₂ ($P = 0.363$ in Fig. 2; $P = 0.344$ in Extended Data Table 1), MAT ($P = 0.789$ in Fig. 2; $P = 0.804$ in Extended Data Table 1) or MCWD ($P = 0.338$ in Fig. 2; $P = 0.325$ in Extended Data Table 1).

We further investigate the responses of carbon gains and losses (for which the above analysis has no explanatory power) by expanding our potential explanatory variables to include five more. These are the changes in environmental conditions (CO₂-change, MAT-change, MCWD-change, see Extended Data Fig. 3 for calculation details) and two attributes of forests that may influence their response to the same environmental changes: the plot mean wood density (which in old-growth forests correlates with belowground resource availability^{28,29}) and the plot carbon residence time (CRT, which measures how long fixed carbon remains in the system and hence reflects when past increases in carbon gains leave the system as elevated carbon losses³⁰).

The minimum adequate carbon gain model using our expanded explanatory variables (best-ranked model using multimodel inference) has a significant positive relationship with CO₂-change ($P = 0.013$), and significant negative relationships with MAT ($P = 0.001$), MAT-change ($P < 0.001$), MCWD ($P < 0.001$) and wood density ($P = 0.015$; Table 2; model-average results are similar, see Methods and Supplementary Tables 2–4). The retention of both MAT and MAT-change suggests that higher temperatures correspond to lower tree growth, and that trees only partially acclimate to recently rising temperatures, which further reduces growth, consistent with warming experiments³¹ and observations⁹. The inclusion of higher wood density and its relationship to lower carbon gains (Extended Data Fig. 4), alongside no significant temporal trends in wood density (Extended Data Fig. 5), suggests that old-growth forests with denser-wooded tree communities typically have fewer available below ground resources, or such patterns may also emerge from disturbance regimes lacking large-scale exogenous events, consistent with previous studies^{26,28,32}.

The minimum adequate carbon gain model using our expanded explanatory variables also highlights continental differences. Between 1 January 2000 and 31 December 2014 modelled African forest carbon

gains increased by 3.1% compared with a 0.1% decline in Amazonia over the same interval (Table 2). In Africa, from 2000 to 2015, the increase in carbon gains was composed of a 3.7% increase from CO₂-change, partially offset by increasing droughts depleting gains by 0.5%, and only a slight decline in gains of 0.1% resulting from temperature increases (Table 2), because the rate of temperature change (MAT-change) decelerated over this time window (Extended Data Fig. 5). For Amazonia, the same 3.7% increase in carbon gains due to CO₂-change was seen. Opposing this trend was increasing droughts—and the greater sensitivity to drought of Amazonian forests—which reduced carbon gains by 2.7% (five times the impact in Africa), and temperature increases at the same rate as in the past (that is, MAT-change is zero) further reduced gains by 1.1% (ten times the impact in Africa), leaving a net change in gains slightly below zero (Table 2). Therefore, the stalling of carbon gain increases in Amazonia in the decade to mid-2011⁶ is a response to drought and temperature and not due to an unexpected saturation of CO₂ fertilization.

Overall, the larger modelled increase in carbon gains in Africa relative to Amazonia appear to be driven by slower warming, fewer or less extreme droughts, lower forest sensitivity to droughts, and overall lower temperatures (African forests are on average ~1.1 °C cooler than Amazonian forests, because they typically grow at higher elevations of ~200 metres above sea level). Other continental differences may also be influencing the results, including higher nitrogen deposition in African tropical forests due to the seasonal burning of nearby savannahs³³ and biogeographical history resulting in differing contemporary species pools and resulting functional attributes^{34,35}.

The minimum adequate carbon loss model using our expanded explanatory variables shows significantly higher losses with CO₂-change ($P = 0.026$) and MAT-change ($P < 0.001$) and significantly lower losses with MCWD ($P = 0.030$) and CRT ($P < 0.001$; Table 2). Thus, changes in carbon losses appear to be largely a function of past carbon gains. First, the greater losses in forests with shorter CRT conform to a ‘high-gain, high-loss’ forest dynamics pattern^{26,28}. Second, wetter plots have a longer growing season and thus they have higher gains and correspondingly higher losses, explaining the negative relationship with MCWD. Third, as increasing CO₂ levels result in additional carbon gains, after some time these additional past gains leave the system, resulting in

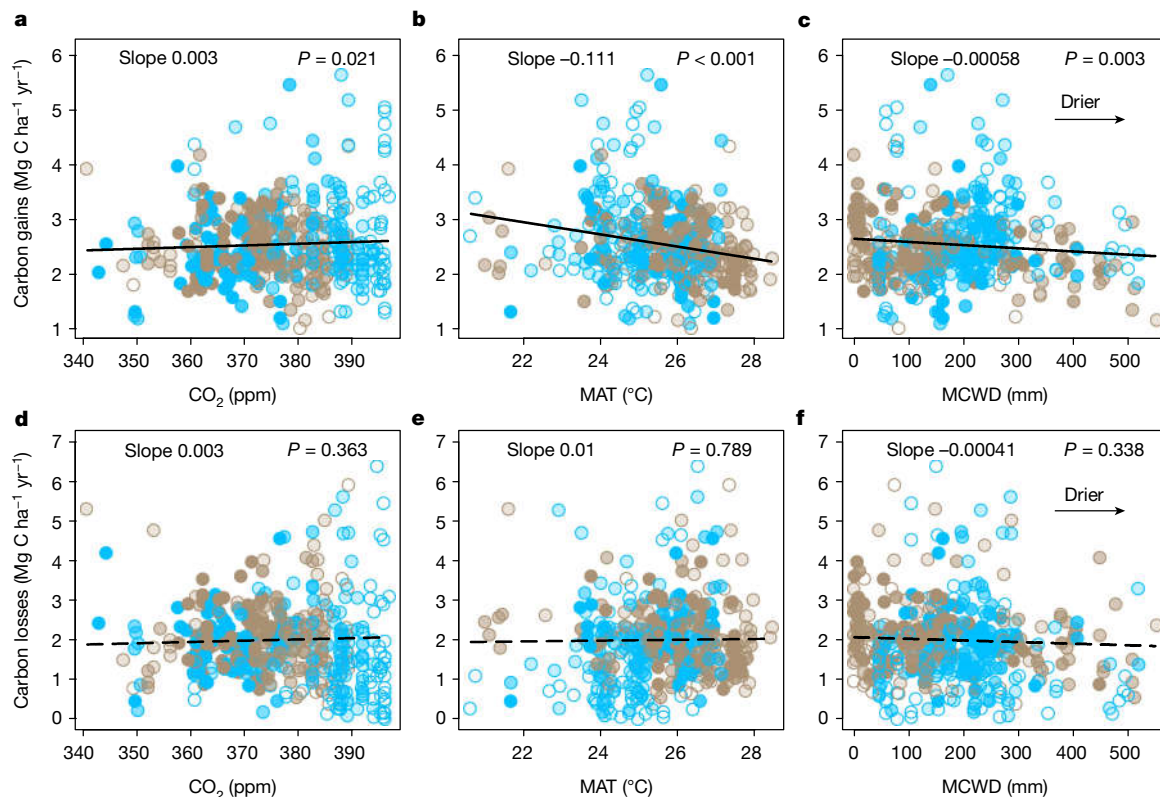


Fig. 2 | Potential environmental drivers of carbon gains and losses in structurally intact old-growth tropical forests in Africa and Amazonia.

Aboveground carbon gains, from woody production (a–c), and aboveground carbon losses, from tree mortality (d–f), are presented as time-weighted mean values for each plot, that is, each census within a plot is weighted by its length, against the corresponding values of atmospheric carbon dioxide concentration (CO₂), temperature (MAT) and drought (MCWD), for African (blue) and Amazonian (brown) inventory plots. For visual clarity each data point therefore represents an inventory plot, and the shading represents the total monitoring length, with empty circles corresponding to plots monitored for ≤5 years and solid circles for plots monitored for >20 years. Solid lines show

significant trends and dashed lines show non-significant trends calculated using linear mixed effects models with census intervals ($n = 1,566$) nested within plots ($n = 565$), using an empirically derived weighting based on interval length and plot area, on the untransformed pooled Africa and Amazon dataset (see Methods). Slopes and P values are from the same linear mixed effects models. Carbon loss data and models are presented untransformed for comparison with carbon gains, but transformation is needed to fit normality assumptions; performing linear mixed effects models on transformed carbon loss data does not change the presented significance trends, nor does including all three parameters and transformed data in a model (see Extended Data Table 1).

greater carbon losses, which explains the positive relationship with CO₂ change. Finally, in addition to these relationships with carbon gains, the inclusion of MAT-change ($P < 0.001$) indicates tree mortality induced by heat or by increased vapour pressure deficit²⁴. Overall, our results imply that chronic long-term environmental change factors, temperature and CO₂, rather than simply the direct effects of drought, underlie longer-term trends in tropical forest tree mortality, although other changes such as rising liana infestation rates seen in Amazonia^{36,37} cannot be excluded.

The minimum adequate carbon loss model using our expanded explanatory variables replicates the continental trends (Fig. 3). The overall modelled lower loss rates in Africa reflect their longer CRT (69 years, 95% CI, 66–72), compared with Amazonian forests (56 years, 95% CI, 54–59) while over the 2000–2015 window the much smaller modelled increase in loss rates in Africa compared to Amazonia results from a slower increase in warming and a stable CRT in Africa compared to continued warming at previous rates and a shortening CRT in Amazonian forests (Table 2; Extended Data Fig. 5). Furthermore, given that losses appear to lag behind gains, they should relate to the long-term CRT of plots. This is what we find: the longer the CRT the smaller the increase in carbon losses, with no increase in losses for plots with CRT ≥ 77 years (Extended Data Fig. 6). Consequently, owing to the typically longer CRT of African forests, increasing losses in Africa ought to appear 10–15 years after the increase in Amazon losses began (around 1995).

Strikingly, in Africa the most intensely monitored plots suggest that losses began increasing from about 2010 (Extended Data Fig. 7), and plots with shorter CRT are driving the increase (Extended Data Fig. 8). Thus, a mortality-dominated decline of the African carbon sink appears to have begun very recently.

Future of the tropical forest carbon sink

Our carbon gain and loss models (Table 2) can be used to make a tentative estimate of the future size of the per unit area intact forest carbon sink (Fig. 3). Extrapolations of the changes in the predictor variables from 1983–2015 forward to 31 December 2039 (Extended Data Fig. 5) show declines in the sink on both continents (Fig. 3). By 2030 the carbon sink in aboveground live biomass in intact African tropical forest is predicted to decline by 14% from the measured 2010–15 mean to 0.57 Mg C ha⁻¹ yr⁻¹ (2σ range, 0.16–0.96; Fig. 3). The Amazon sink continues to rapidly decline, reaching zero in 2035 (2σ range, 2011–2089; Fig. 3). Our estimated sink strength on both continents in the 2020s and 2030s is sensitive to future CO₂ emissions pathways (CO₂-change)³⁸, resulting temperature increase (MAT, MAT-change) and hydrological changes (MCWD), plus changes in forest dynamics (CRT), but the sink is always lower than levels seen in the 2000s (see Methods and Supplementary Table 5). Therefore, the carbon sink strength of the world's two most extensive tropical forests have now saturated, albeit asynchronously.

Table 2 | Minimum adequate models to predict carbon gains and losses in African and Amazonian forests

Carbon gains (Mg C ha ⁻¹ yr ⁻¹)					
Predictor variable	Parameter value	Standard error	t value	P value	2000–2015 change in gains (%) ^a
Intercept	5.255 5.395	0.603 0.614	8.7 8.8	<0.001	–
CO ₂ -change (ppm yr ⁻¹) ^b	0.238	0.096	2.5	0.013	3.69% 3.71%
MAT (°C)	–0.083	0.025	–3.3	0.001	–0.67% –1.07%
MAT-change (°C yr ⁻¹) ^c	–1.243	0.233	–5.3	<0.001	0.58% 0.00% ^d
MCWD (mm × 1,000)	–0.405 –1.391	0.381 0.24	–1.1 –5.8	0.289 <0.001	–0.52% –2.73%
Wood density (g cm ⁻³)	–1.295	0.530	–2.4	0.015	0.05% 0.00%
Carbon losses (Mg C ha ⁻¹ yr ⁻¹) ^e					
Predictor variable	Parameter value	Standard error	t value	P value	2000–2015 change in losses (%) ^a
Intercept	1.216	0.086	14.1	<0.001	–
CO ₂ -change (ppm yr ⁻¹) ^b	0.130	0.059	2.2	0.026	11.38% 14.81%
MAT-change (°C yr ⁻¹) ^c	0.766	0.162	4.7	<0.001	–1.56% 0.00%
MCWD (mm × 1,000)	–0.232	0.107	–2.2	0.030	–1.21% –2.42%
CRT (years)	–0.003	0.001	–6.1	<0.001	–0.57% 1.39%

This table shows the best-ranked gains and loss models. Where continental values differ, those for Africa are reported first, followed by ‘|’, then the Amazonian values.
^aThe 1 January 2000 to 31 December 2014 change in gains/losses for each predictor variable was estimated allowing only the focal predictor to vary; this change was then expressed as a percentage of the annual gains/losses in the year 2000, allowing all predictors to vary.
^bChange over the past 56 years (see Extended Data Fig. 3).
^cChange over the past 5 years (see Extended Data Fig. 3).
^dA positive value for Africa indicates that MAT increased more slowly over 2000–2015 compared to the mean increase over 1983–2015, therefore contributing to an increase in gains; a zero value for Amazonia indicates that the rate of MAT increase was the same over 2000–2015 as the mean increase over 1983–2015.
^eCarbon loss values were normalized via power-law transformation, with power parameter $\lambda = 0.361$.

Scaling results to the pan-tropics

Scaling our estimated mean sink strength by forest area for each continent signifies that Earth has passed the point of peak carbon sequestration into intact tropical forests (Table 1). The continental sink in Amazonia peaked in the 1990s, followed by a decline, driven by sink strength peaking in the 1990s and a continued decline in forest area (Table 1). In Africa the per unit area sink strength peaked later, in the 2000–2010 period, but the continental African sink peaked in the 1990s, owing to the decline in forest area in the 2000s outpacing the small per unit area increase in sink strength. Including the modest uptake in the much smaller area of intact Asian tropical forest¹⁵ indicates that total pan-tropical carbon uptake peaked in the 1990s (Table 1). From the peak pan-tropical intact forest uptake of 1.26 Pg C yr⁻¹ in the 1990s, we project a continued decline reaching just 0.29 Pg C yr⁻¹ in the 2030s (multi-decade decline of –0.24 Pg C yr⁻¹ per decade), driven by (1) reduced mean pan-tropical sink strength decline of 0.1 Mg C ha⁻¹ yr⁻¹ per decade and (2) ongoing forest area losses of –13.5 million ha yr⁻¹ (see Extended Data Table 2 for forest area details). Critically, climate-driven vegetation model simulations have not predicted that the peak net carbon uptake into intact tropical forests has already been passed^{2,4,5}.

Discussion

Our method of scaling to arrive at a pan-tropical sink estimate—in common with other studies using similar datasets^{1,6,13}—is limited. Yet, pervasive net carbon uptake is expected given that we find a strong and ongoing CO₂ fertilization effect. Using our CO₂ response in Table 2, we find an increase in aboveground carbon stocks of 10.8 ± 3.7 Mg C ha⁻¹ per 100 ppm CO₂, equivalent to 6.5 ± 2.2% (±standard error; using an area-weighted pan-tropical mean aboveground carbon stock of 165 Mg C ha⁻¹). This is comparable to the 5.0 ± 1.2% increase in tropical forest C stocks per 100 ppm CO₂ derived from a recent synthesis of CO₂ fertilization experiments, despite a lack of data from old-growth tropical forests³⁹. Our result is within the range of climate-driven vegetation models^{2,7}, although it is greater than results from a number of recently published models that include potential nutrient constraints, reported as 5.9 ± 4.7 Mg C ha⁻¹ per 100 ppm CO₂ (ref. ⁴⁰). We find that

the CO₂ fertilization-driven uptake is currently only partially offset by the negative impacts of similarly widespread rising air temperatures (–2.0 ± 0.4 Mg C ha⁻¹ °C⁻¹, from Table 2), consistent with models⁷, limited experiments³¹ and independent observations⁹, plus well-documented negative responses to drought^{41,42}. Long-term and extensive increases in satellite-derived greenness in tropical regions that have not experienced major changes in land-use management^{17,43}, particularly in central Africa in the past decade⁴⁴, indicate increases in tropical forest net primary productivity, providing further evidence that the sink is a widespread phenomenon.

Nonetheless, our analyses suggest that this pervasive intact tropical forest sink in live biomass is in long-term decline, having peaked first in Amazonia, and more recently followed by African forests, explaining the prior Africa–Amazon carbon sink divergence as part of a longer-term pattern of asynchronous saturation and decline. Over time, the continued CO₂ fertilization effect is being increasingly counteracted by the impacts of higher temperatures and droughts on tree growth and mortality, which are modulated by internal forest dynamics, with forests with the shortest CRT saturating first.

From an atmospheric perspective, the full impacts of the contribution to the saturation of the sink from slowing carbon gains are experienced immediately, but the contribution from rising carbon losses is delayed because dead trees do not decompose instantaneously. Decomposition of this dead tree mass is about half complete in 4 years, and about 85% complete in 10 years, so rising carbon losses result in delayed carbon additions to the atmosphere⁴⁵. Hence, from an atmospheric perspective, the intact tropical forest biomass carbon sink probably peaked a few years later than our inventory data indicate and the full impacts are not yet realized. The pan-tropical carbon sink in live biomass declined by 0.27 Pg C yr⁻¹ between the 1990s and 2000s (Table 1), but accounting for dead wood decomposition⁴⁵ shows a smaller 0.17 Pg C yr⁻¹ reduction from an atmospheric perspective (see Methods).

Given that the overall global terrestrial carbon sink is increasing, a weakening intact tropical forest sink implies that the extra-tropical carbon sink has increased over the past two decades. Independent observations of interhemispheric atmospheric CO₂ concentration

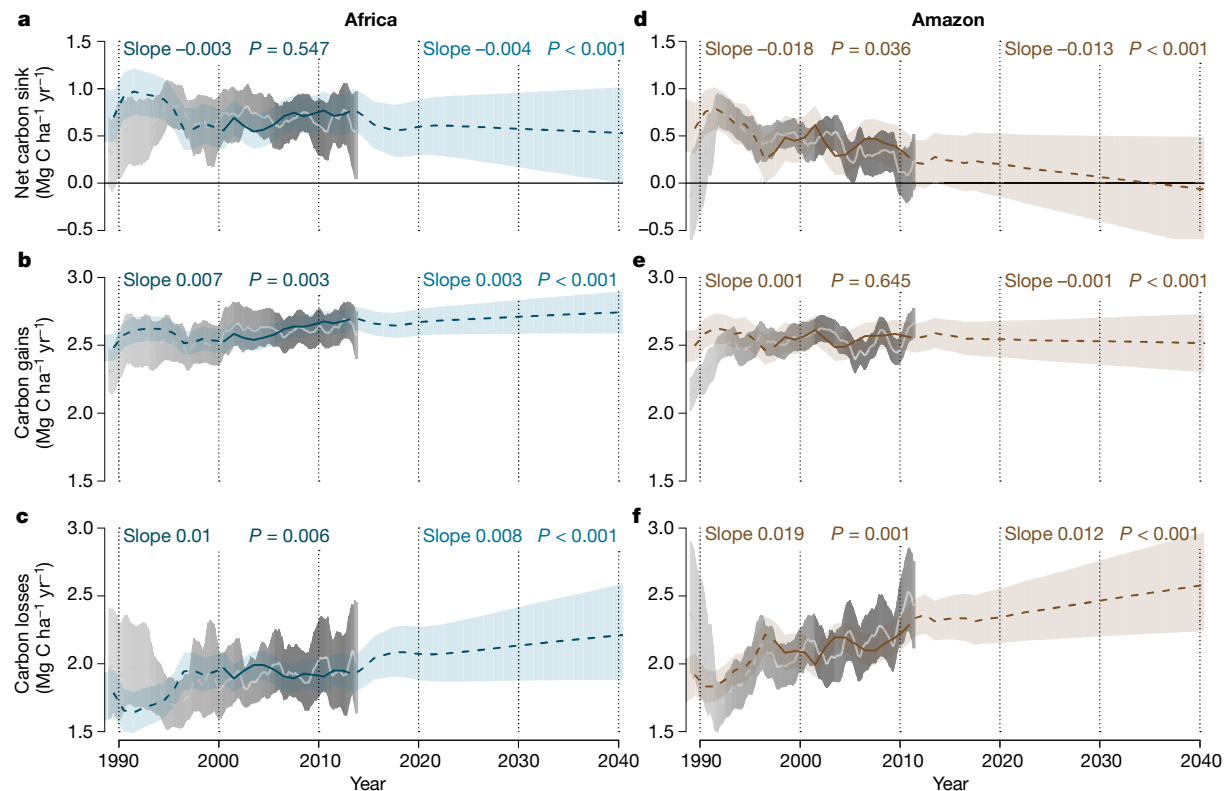


Fig. 3 | Modelled past and future carbon dynamics of structurally intact old-growth tropical forests in Africa and Amazonia. **a–f**, Predictions of net aboveground live biomass carbon (**a, d**), carbon gains (**b, e**), and carbon losses (**c, f**), for African (left panels) and Amazonian (right panels) plot inventory networks, based on CO_2 -change, MAT, MAT-change, drought (MCWD), plot wood density, and plot CRT, using observations in Africa until 31 December 2014 and Amazonia until mid-2011, and extrapolations of prior trends to

31 December 2039. Model predictions are in blue (Africa) and brown (Amazon), with solid lines spanning the window when $\geq 75\%$ of plots were monitored to show model consistency with the observed trends, and shading showing upper and lower confidence intervals accounting for uncertainties in the model (both fixed and random effects) and uncertainties in the predictor variables. Light-grey lines and grey shading are the mean and 95% CI of the observations from the African and Amazonian plot networks.

indicates that carbon uptake into the Northern Hemisphere landmass has increased at a greater rate than the global terrestrial carbon sink since the 1990s, with a further disproportionate increase in the 2000s¹⁰. The interhemispheric analysis suggests a weakening of the tropical forest sink by $-0.2 \text{ Pg C yr}^{-1}$ between the 1990s and 2000s¹⁰, which is similar to the $0.17 \text{ Pg C yr}^{-1}$ weakening over the same time period that we find. This reinforces our conclusion that the intact tropical forest carbon sink has already saturated.

In summary, our results indicate that although intact tropical forests remain major stores of carbon and are key centres of biodiversity¹¹, their ability to sequester additional carbon in trees is waning. In the 1990s intact tropical forests removed 17% of anthropogenic CO_2 emissions. This declined to an estimated 6% in the 2010s, because the pan-tropical weighted average per unit area sink strength declined by 33%, forest area decreased by 19% and anthropogenic CO_2 emissions increased by 46%. Although tropical forests are more immediately threatened by deforestation⁴⁶ and degradation⁴⁷, and the future carbon balance will also depend on secondary forest dynamics⁴⁸ and forest restoration plans⁴⁹, our analyses show that they are also affected by atmospheric chemistry and climatic changes. Given that the intact tropical forest carbon sink is set to end sooner than even the most pessimistic climate-driven vegetation models predict^{4,5}, our analyses suggest that climate change impacts in the tropics may become more severe than predicted. Furthermore, the carbon balance of intact tropical forests will only stabilize once CO_2 concentrations and the climate stabilizes.

Continued on-the-ground monitoring of the world's remaining intact tropical forests will be required to test our prediction that the carbon sink in live trees will continue to decline, particularly as future changes in the tree species composition may alter the resilience of the sink and

because we cannot exclude the possibility of decadal-scale climate impacts on these forests. Such direct ground-based measurements also provide a constraint on estimating the size, location and climate sensitivity of the terrestrial carbon sink. In addition, our conclusion that tree mortality and internal forest dynamics are important controls on the future of the tropical forest carbon sink may assist in improving the vegetation components of Earth System Models⁵⁰ and contribute to reducing terrestrial carbon cycle feedback uncertainty^{19,20}. Our findings also have policy implications. At the individual country level, given that intact tropical forests are a carbon sink but the rate of reduction will differ continentally and probably regionally (for example, aseasonal Amazon forests are less affected by droughts), national greenhouse gas reporting will require careful forest monitoring. At the international level, given that tropical forests are likely to sequester less carbon in the future than Earth System Models predict, an earlier date by which to reach net zero anthropogenic greenhouse gas emissions will be required to meet any given commitment to limit the global heating of Earth.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2035-0>.

1. Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
2. Stith, S. et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences* **12**, 653–679 (2015).

3. Gaubert, B. et al. Global atmospheric CO₂ inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate. *Biogeosciences* **16**, 117–134 (2019).
4. Huntingford, C. et al. Simulated resilience of tropical rainforests to CO₂-induced climate change. *Nat. Geosci.* **6**, 268–273 (2013).
5. Mercado, L. M. et al. Large sensitivity in land carbon storage due to geographical and temporal variation in the thermal response of photosynthetic capacity. *New Phytol.* **218**, 1462–1477 (2018).
6. Brienen, R. J. W. et al. Long-term decline of the Amazon carbon sink. *Nature* **519**, 344–348 (2015).
7. Piao, S. et al. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO₂ trends. *Glob. Change Biol.* **19**, 2117–2132 (2013).
8. Schimel, D., Stephens, B. B. & Fisher, J. B. Effect of increasing CO₂ on the terrestrial carbon cycle. *Proc. Natl Acad. Sci. USA* **112**, 436–441 (2015).
9. Anderegg, W. R. L. et al. Tropical nighttime warming as a dominant driver of variability in the terrestrial carbon sink. *Proc. Natl Acad. Sci. USA* **112**, 15591–15596 (2015).
10. Ciais, P. et al. Five decades of northern land carbon uptake revealed by the interhemispheric CO₂ gradient. *Nature* **568**, 221–225 (2019).
11. Lewis, S. L., Edwards, D. P. & Galbraith, D. Increasing human dominance of tropical forests. *Science* **349**, 827–832 (2015).
12. Pugh, T. A. M. et al. Role of forest regrowth in global carbon sink dynamics. *Proc. Natl Acad. Sci. USA* **116**, 4382–4387 (2019).
13. Lewis, S. L. et al. Increasing carbon storage in intact African tropical forests. *Nature* **457**, 1003–1006 (2009).
14. Phillips, O. L. et al. Drought sensitivity of the Amazon rainforest. *Science* **323**, 1344–1347 (2009).
15. Qie, L. et al. Long-term carbon sink in Borneo's forests halted by drought and vulnerable to edge effects. *Nat. Commun.* **8**, 1966 (2017).
16. Gatti, L. V. et al. Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements. *Nature* **506**, 76–80 (2014).
17. Nemani, R. R. et al. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
18. Keenan, T. F. et al. Recent pause in the growth rate of atmospheric CO₂ due to enhanced terrestrial carbon uptake. *Nat. Commun.* **7**, 13428 (2016).
19. Booth, B. B. B. et al. High sensitivity of future global warming to land carbon cycle processes. *Environ. Res. Lett.* **7**, 024002 (2012).
20. Lombardozzi, D. L., Bonan, G. B., Smith, N. G., Dukes, J. S. & Fisher, R. A. Temperature acclimation of photosynthesis and respiration: a key uncertainty in carbon cycle-climate feedback. *Geophys. Res. Lett.* **42**, 8624–8631 (2015).
21. Le Quéré, C. et al. Global carbon budget 2018. *Earth Syst. Sci. Data* **10**, 2141–2194 (2018).
22. Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M. F. & Nepstad, D. The 2010 Amazon drought. *Science* **331**, 554 (2011).
23. Feldpausch, T. R. et al. Amazon forest response to repeated droughts. *Glob. Biogeochem. Cycles* **30**, 964–982 (2016).
24. McDowell, N. et al. Drivers and mechanisms of tree mortality in moist tropical forests. *New Phytol.* **219**, 851–869 (2018).
25. Aleixo, I. et al. Amazonian rainforest tree mortality driven by climate and functional traits. *Nat. Clim. Chang.* **9**, 384–388 (2019).
26. Lewis, S. L. et al. Concerted changes in tropical forest structure and dynamics: evidence from 50 South American long-term plots. *Phil. Trans. R. Soc. Lond. B* **359**, 421–436 (2004).
27. Lewis, S. L. et al. Above-ground biomass and structure of 260 African tropical forests. *Phil. Trans. R. Soc. Lond. B* **368**, 2010295 (2013).
28. Quesada, C. A. et al. Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences* **9**, 2203–2246 (2012).
29. Malhi, Y. et al. The above-ground coarse wood productivity of 104 neotropical forest plots. *Glob. Change Biol.* **10**, 563–591 (2004).
30. Galbraith, D. et al. Residence times of woody biomass in tropical forests. *Plant Ecol. Divers.* **6**, 139–157 (2013).
31. Reich, P. B. et al. Boreal and temperate trees show strong acclimation of respiration to warming. *Nature* **531**, 633–636 (2016).
32. ter Steege, H. et al. Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* **443**, 444–447 (2006).
33. Batters, M. et al. High fire-derived nitrogen deposition on central African forests. *Proc. Natl Acad. Sci. USA* **115**, 549–554 (2018).
34. Parmentier, I. et al. The odd man out? Might climate explain the lower tree alpha-diversity of African rain forests relative to Amazonian rain forests? *J. Ecol.* **95**, 1058–1071 (2007).
35. Slik, J. W. F. et al. Phylogenetic classification of the world's tropical forests. *Proc. Natl Acad. Sci. USA* **115**, 1837–1842 (2018).
36. Phillips, O. L. et al. Increasing dominance of large lianas in Amazonian forests. *Nature* **418**, 770–774 (2002).
37. Schnitzer, S. A. & Bongers, F. Increasing liana abundance and biomass in tropical forests: emerging patterns and putative mechanisms. *Ecol. Lett.* **14**, 397–406 (2011).
38. Meinshausen, M. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).
39. Terrer, C. et al. Nitrogen and phosphorus constrain the CO₂ fertilization of global plant biomass. *Nat. Clim. Chang.* **9**, 684–689 (2019).
40. Fleischer, K. et al. Amazon forest response to CO₂ fertilization dependent on plant phosphorus acquisition. *Nat. Geosci.* **12**, 736–741 (2019).
41. Jiang, Y. et al. Widespread increase of boreal summer dry season length over the Congo rainforest. *Nat. Clim. Chang.* **9**, 617–622 (2019).
42. Gloor, M. et al. Recent Amazon climate as background for possible ongoing and future changes of Amazon humid forests. *Glob. Biogeochem. Cycles* **29**, 1384–1399 (2015).
43. Kolby Smith, W. et al. Large divergence of satellite and Earth system model estimates of global terrestrial CO₂ fertilization. *Nat. Clim. Chang.* **6**, 306–310 (2016).
44. Chen, C. et al. China and India lead in greening of the world through land-use management. *Nature Sustain.* **2**, 122–129 (2019).
45. Chambers, J. Q., Higuchi, N., Schimel, J. P., Ferreira, L. V. & Melack, J. M. Decomposition and carbon cycling of dead trees in tropical forests of the central Amazon. *Oecologia* **122**, 380–388 (2000).
46. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
47. Pearson, T. R. H., Brown, S., Murray, L. & Sidman, G. Greenhouse gas emissions from tropical forest degradation: an underestimated source. *Carbon Balance Manag.* **12**, 3 (2017).
48. Schwartz, N. B., Uriarte, M., DeFries, R., Gutierrez-Velez, V. H. & Pinedo-Vasquez, M. A. Land-use dynamics influence estimates of carbon sequestration potential in tropical second-growth forest. *Environ. Res. Lett.* **12**, 074023 (2017).
49. Lewis, S. L., Wheeler, C. E., Mitchard, E. T. A. & Koch, A. Regenerate natural forests to store carbon. *Nature* **568**, 25–28 (2019).
50. Yu, K. et al. Pervasive decreases in living vegetation carbon turnover time across forest climate zones. *Proc. Natl Acad. Sci. USA* **116**, 24662–24667 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Wannes Hubau^{1,2,3,95,52}, Simon L. Lewis^{1,4,95}, Oliver L. Phillips¹, Kofi Affum-Baffoe⁵, Hans Beekman², Aida Cuní-Sánchez^{4,6}, Armandu K. Daniels⁷, Corneille E. N. Ewango^{8,9,10}, Sophie Fauset¹¹, Jacques M. Mukinzi^{8,12,13}, Douglas Sheil¹⁴, Bonaventure Sonké¹⁵, Martin J. P. Sullivan¹¹⁶, Terry C. H. Sunderland^{17,18}, Hermann Taedoum^{15,19}, Sean C. Thomas²⁰, Lee J. T. White^{21,22,23}, Katharine A. Abernethy^{22,23}, Stephen Adu-Bredu²⁴, Christian A. Amani^{17,25}, Timothy R. Baker¹, Lindsay F. Banin²⁶, Fidèle Baya^{27,28}, Serge K. Begne¹³⁵, Amy C. Bennett¹, Fabrice Benedet^{29,30}, Robert Bitariho³¹, Yannick E. Bocko³², Pascal Boeckx³³, Patrick Boundja^{37,34}, Roel J. W. Brienen¹, Terry Brncic³⁴, Eric Chezeaux³⁵, George B. Chuyong³⁶, Connie J. Clark³⁷, Murray Collins^{38,39}, James A. Comiskey^{40,41}, David A. Coomes⁴², Greta C. Dargie¹, Thales de Haulleville², Marie Noel Djoukouo Kamdem³⁶, Jean-Louis Doucet⁴³, Adriane Esquivel-Muelbert^{1,44}, Ted R. Feldpausch⁴⁵, Alusine Fofanah⁴⁶, Ernest G. Foli²⁴, Martin Gilpin¹, Emanuel Gloor¹, Christelle Gnomadje⁴⁷, Sylvie Gourlet-Fleury^{29,30}, Jefferson S. Hall¹⁸, Alan C. Hamilton⁴⁹, David J. Harris⁵⁰, Terese B. Hart^{51,52}, Mireille B. N. Hockembia³⁴, Annette Hladik⁵³, Suspense A. Ife⁵⁴, Kathryn J. Jeffery²³, Tommaso Jucker⁵⁵, Emmanuel Kasongo Yakusu^{2,3,10}, Elizabeth Kearsley^{2,56}, David Kenfack^{48,57}, Alexander Koch^{1,58}, Miguel E. Leal⁵⁹, Aurora Levesley¹, Jeremy A. Lindsell^{60,61}, Janvier Lisingo⁶², Gabriela Lopez-Gonzalez¹, Jon C. Lovett^{1,63}, Jean-Remy Makana⁶², Yadvinder Malhi⁶⁴, Andrew R. Marshall^{6,65,66}, Jim Martin⁶⁷, Emanuel H. Martin^{57,68}, Faustina M. Mbayu¹⁰, Vincent P. Medjibe^{37,69,70}, Vianet Mihindou^{21,70}, Edward T. A. Mitchard³⁸, Sam Moore⁶⁸, Pantaleo K. T. Munishi⁷¹, Natacha Nssi Bengone²¹, Lucas Ojo⁷², Fidèle Evouna Ondo⁷⁰, Kelvin S.-H. Peh^{73,74}, Georgia C. Pickavance¹, Axel Dalberg Poulsen⁵⁰, John R. Poulsen³⁷, Lan Qie^{1,75}, Jan Reitsma⁷⁶, Francesco Rovero^{77,78}, Michael D. Swaine⁷⁹, Joey Talbot^{1,80}, James Taplin⁸¹, David M. Taylor⁸², Duncan W. Thomas⁸³, Benjamin Toirambe^{2,84}, John Tshibamba Mukendi^{2,10,85}, Darlington Tuagben^{7,86}, Peter M. Umunay^{87,88}, Geertje M. F. van der Heijden⁸⁹, Hans Verbeeck⁵⁰, Jason Vleminckx^{90,91}, Simon Willcock⁹², Hannsjörg Wöll⁹³, John T. Woods⁹⁴ & Lise Zemagho¹⁵

¹School of Geography, University of Leeds, Leeds, UK. ²Service of Wood Biology, Royal Museum for Central Africa, Tervuren, Belgium. ³Department of Environment, Laboratory of Wood Technology (Woodlab), Ghent University, Ghent, Belgium. ⁴Department of Geography, University College London, London, UK. ⁵Mensuration Unit, Forestry Commission of Ghana, Kumasi, Ghana. ⁶Department of Environment and Geography, University of York, York, UK. ⁷Forestry Development Authority of the Government of Liberia (FDA), Monrovia, Liberia. ⁸DR Congo Programme, Wildlife Conservation Society, Kinshasa, Democratic Republic of Congo. ⁹Centre de Formation et de Recherche en Conservation Forestière (CEFRECOCF), Epulu, Democratic Republic of Congo. ¹⁰Faculté de Gestion de Ressources Naturelles Renouvelables, Université de Kisangani, Kisangani, Democratic Republic of Congo. ¹¹School of Geography, Earth and Environmental Sciences, University of Plymouth, Plymouth, UK. ¹²Salonga National Park, Kinshasa, Democratic Republic of Congo. ¹³World Wide Fund for Nature, Gland, Switzerland. ¹⁴Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway. ¹⁵Plant Systematic and Ecology Laboratory, Higher Teachers' Training College, University of Yaounde I, Yaounde, Cameroon. ¹⁶Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK. ¹⁷Center for International Forestry Research (CIFOR), Bogor, Indonesia. ¹⁸Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada. ¹⁹Bioversity International, Yaounde, Cameroon. ²⁰Faculty of Forestry, University of Toronto, Toronto, Ontario, Canada. ²¹Ministry of Forests, Seas, Environment and Climate, Libreville, Gabon. ²²Institut de Recherche en Écologie Tropicale, Libreville, Gabon. ²³Department of Biological and Environmental Sciences, University of Stirling, Stirling, UK. ²⁴Forestry Research Institute of Ghana (FORIG), Kumasi, Ghana. ²⁵Université Officielle de Bukavu, Bukavu, Democratic Republic of Congo. ²⁶UK Centre for Ecology & Hydrology, Penicuik, UK. ²⁷Ministère des Eaux, Forêts, Chasse et Pêche (MEFCP), Bangui, Central African Republic. ²⁸Institut Centrafricain de Recherche Agronomique (ICRA), Bangui, Central African Republic. ²⁹Forêts et Sociétés (F&S), Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France. ³⁰Forêts et Sociétés (F&S), Université de Montpellier, Montpellier, France. ³¹The Institute of Tropical Forest Conservation (ITFC), Mbarara University of Science and Technology (MUST), Mbarara, Uganda. ³²Faculté des

Sciences et Techniques, Laboratoire de Botanique et Écologie, Université Marien Ngouabi, Brazzaville, Republic of Congo. ³³Isotope Bioscience Laboratory-ISOFYS, Ghent University, Ghent, Belgium. ³⁴Congo Programme, Wildlife Conservation Society, Brazzaville, Republic of Congo. ³⁵Rougier-Gabon, Libreville, Gabon. ³⁶Faculty of Science, Department of Botany and Plant Physiology, University of Buea, Buea, Cameroon. ³⁷Nicholas School of the Environment, Duke University, Durham, NC, USA. ³⁸School of GeoSciences, University of Edinburgh, Edinburgh, UK. ³⁹Grantham Research Institute on Climate Change and the Environment, London, UK. ⁴⁰Inventory and Monitoring Program, National Park Service, Fredericksburg, VA, USA. ⁴¹Smithsonian Institution, Washington, DC, USA. ⁴²Department of Plant Sciences, University of Cambridge, Cambridge, UK. ⁴³TERRA, Forest is Life, Gembloux Agro-Bio Tech, University of Liège, Liège, Belgium. ⁴⁴School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK. ⁴⁵Department of Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, UK. ⁴⁶The Gola Rainforest National Park, Kenema, Sierra Leone. ⁴⁷National Herbarium, Yaounde, Cameroon. ⁴⁸Forest Global Earth Observatory (ForestGEO), Smithsonian Tropical Research Institute, Washington, DC, USA. ⁴⁹Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China. ⁵⁰Royal Botanic Garden Edinburgh, Edinburgh, UK. ⁵¹Lukuru Wildlife Research Foundation, Kinshasa, Democratic Republic of Congo. ⁵²Division of Vertebrate Zoology, Yale Peabody Museum of Natural History, New Haven, CT, USA. ⁵³Département Hommes et Environnement, Muséum National d'Histoire Naturel, Paris, France. ⁵⁴École Normale Supérieure (ENS), Département des Sciences et Vie de la Terre, Laboratoire de Géomatique et d'Écologie Tropicale Appliquée, Université Marien Ngouabi, Brazzaville, Republic of Congo. ⁵⁵School of Biological Sciences, University of Bristol, Bristol, UK. ⁵⁶Department of Environment, Laboratory of Computational & Applied Vegetation Ecology (Cavelab), Ghent University, Ghent, Belgium. ⁵⁷Tropical Ecology, Assessment and Monitoring (TEAM) Network, Arlington, VA, USA. ⁵⁸Department of Earth Sciences, University of Hong Kong, Hong Kong, China. ⁵⁹Uganda Programme, Wildlife Conservation Society, Kampala, Uganda. ⁶⁰A Rocha International, Cambridge, UK. ⁶¹Centre for Conservation Science, The Royal Society for the Protection of Birds, Sandy, UK. ⁶²Faculté des Sciences, Laboratoire d'Écologie et Aménagement Forestier, Université de Kisangani, Kisangani, Democratic Republic of Congo. ⁶³Royal Botanic Gardens, Kew, London, UK. ⁶⁴Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK. ⁶⁵Tropical Forests and People Research Centre, University of the Sunshine Coast, Sippy Downs, Queensland, Australia. ⁶⁶Flamingo Land Ltd, Kirby Misperton, UK. ⁶⁷Fleming College, Peterborough, Ontario, Canada. ⁶⁸Udzungwa Ecological Monitoring Centre, Mang'ula, Tanzania. ⁶⁹Commission of Central African Forests (COMIFAC), Yaounde, Cameroon. ⁷⁰Agence Nationale des Parcs Nationaux, Libreville, Gabon. ⁷¹Sokoine University of Agriculture, Morogoro, Tanzania. ⁷²University of Abeokuta, Abeokuta, Nigeria. ⁷³School of Biological Sciences, University of Southampton, Southampton, UK. ⁷⁴Department of Zoology, Conservation Science Group, University of Cambridge, Cambridge, UK. ⁷⁵School of Life Sciences, University of Lincoln, Lincoln, UK. ⁷⁶Bureau Waardenburg, Culemborg, The Netherlands. ⁷⁷Department of Biology, University of Florence, Florence, Italy. ⁷⁸Tropical Biodiversity Section, MUSE—Museo delle Scienze, Trento, Italy. ⁷⁹Department of Plant & Soil Science, School of Biological Sciences, University of Aberdeen, Aberdeen, UK. ⁸⁰Institute for Transport Studies, University of Leeds, Leeds, UK. ⁸¹UK Research & Innovation, Innovate UK, London, UK. ⁸²Department of Geography, National University of Singapore, Singapore, Singapore. ⁸³Biology Department, Washington State University, Vancouver, WA, USA. ⁸⁴Ministère de l'Environnement et Développement Durable, Kinshasa, Democratic Republic of Congo. ⁸⁵Faculté des Sciences Appliquées, Université de Mbujimayi, Mbujimayi, Democratic Republic of Congo. ⁸⁶Friends of Ecosystem and the Environment, Monrovia, Liberia. ⁸⁷Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA. ⁸⁸Wildlife Conservation Society, New York, NY, USA. ⁸⁹School of Geography, University of Nottingham, Nottingham, UK. ⁹⁰International Center for Tropical Botany, Department of Biological Sciences, Florida International University, Miami, FL, USA. ⁹¹Faculté des Sciences, Service d'Évolution Biologique et Écologie, Université Libre de Bruxelles, Brussels, Belgium. ⁹²School of Natural Sciences, University of Bangor, Bangor, UK. ⁹³Independent Researcher, Bad Aussee, Austria. ⁹⁴W.R.T. College of Agriculture and Forestry, University of Liberia, Monrovia, Liberia. ⁹⁵These authors contributed equally: Wannes Hubau, Simon L. Lewis. [✉]e-mail: whubau@gmail.com

Plot selection

Closed canopy (that is, not woody savannah) old-growth mixed-age forest inventory plots were selected using commonly used criteria^{6,13,27}: structurally intact (that is, free of fire and industrial logging); all trees with diameter at reference height ≥ 100 mm measured at least twice; area ≥ 0.2 ha; altitude $< 1,500$ m above sea level; MAT ≥ 20.0 °C⁵¹; annual precipitation $\geq 1,000$ mm⁵¹; located ≥ 50 m from anthropogenic forest edges. Of the 244 plots included in the study, 217 contribute to the African Tropical Rainforest Observatory Network (AfriTRON; www.afritron.org), with data curated at www.ForestPlots.net^{52,53}. These include plots from Sierra Leone, Liberia, Ghana, Nigeria, Cameroon, Gabon, Republic of Congo, Democratic Republic of Congo, Uganda and Tanzania^{52,53} (Extended Data Fig. 1). Fifteen plots are part of the TEAM network, from Cameroon, Republic of Congo, Tanzania and Uganda^{54–57}. Nine plots contribute to the ForestGEO network, from Cameroon and Democratic Republic of Congo⁵⁸ (9 plots from Democratic Republic of Congo, with codes SNG, contribute to both AfriTRON and ForestGEO networks, included above in the AfriTRON total). Finally, three plots from Central African Republic are part of the CIRAD network^{59,60}. The large majority of plots are sited in terra firme (not inundated by river water) forests and have mixed species composition, although four are in seasonally flooded forest and 14 plots are in *Gilbertiodendron dewevrei* monodominant forest, a locally common forest type in Africa (Supplementary Table 1). The 244 plots have a mean size of 1.1 ha (median, 1 ha), with a total plot area of 277.9 ha. The dataset comprises 391,968 diameter measurements on 135,625 stems, of which 89.9% were identified to species, 97.5% to genus and 97.8% to family. Mean total monitoring period is 11.8 years, mean census length 5.7 years, with a total of 3,214 hectare years of monitoring. The 321 Amazon plots are published and were selected using the same criteria⁶, except in the African selection criteria we specified a minimum anthropogenic edge distance and added a minimum temperature threshold.

Plot inventory and tree biomass carbon estimation

Tree-level aboveground biomass carbon is estimated using an allometric equation with parameters for tree diameter (in mm), tree height (in m) and wood mass density (in g cm⁻³)⁶¹. The calculation of each is discussed in turn. All calculations were performed using the R statistical platform, version 3.2.1 (ref. ⁶²) using the BiomasaFP R package, version 0.2.1 (ref. ⁶³).

Tree diameter. In all plots, all woody stems with ≥ 100 mm diameter at 1.3 m from the base of the stem ('diameter at breast height', DBH, in mm), or 0.5 m above deformities or buttresses, were measured, mapped and identified using standard forest inventory methods^{64,65}. The height of the point of measurement (POM) was marked on the trees and recorded, so that the same POM is used at the subsequent forest census. For stems developing deformities or buttresses over time that could potentially disturb the initial POM, the POM was raised approximately 500 mm above the deformity. Estimates of the diameter growth of trees with changed POM used the ratio of new to old POMs, to create a single trajectory of growth from the series of diameters at two POM heights^{6,13,65}. We used standardized protocols to assess typographical errors and potentially erroneous diameter values (for example, trees shrinking by > 5 mm), missing values, failures to find the original POM, and other issues. Where necessary we estimated the likely value via interpolation or extrapolation from other measurements of that tree, or when this was not possible we used the median or mean growth rate of trees in the same plot, census and size-class. We used the median growth rate for size classes of DBH = 100–199 mm and 200–399 mm. We used the mean growth rate for a size class with DBH > 400 mm, as there were fewer trees in the largest size class⁶⁵. We interpolated measurements for 1.3% of diameters, extrapolated 0.9%, and used median growth rates for 1.5%.

Tree height. Height of individuals from ground to the top leaf, hereafter H_t , was measured in 204 plots, using a laser hypsometer (Nikon forestry Pro) from directly below the crown (most plots), a laser or ultrasonic distance device with an electronic tilt sensor, a manual clinometer, or by direct measurement, that is, climbing the tree. Only trees where the top was visible were selected⁶⁶. In most plots, tree selection was similar: the 10 largest trees were measured, together with 10 randomly selected trees per diameter from five classes: 100–199 mm, 200–299 mm, 300–399 mm, 400–499 mm, and 500+ mm trees, following standard protocols⁶⁶. We measured the actual height of 24,270 individual trees from 204 plots. We used these data and the `local.heights` function in R package `BiomasaFP`⁶³ to fit 3-parameter Weibull relationships:

$$H_t = a(1 - e^{-b\text{DBH}^c}) \quad (1)$$

We chose the Weibull model (with Weibull parameters a , b and c) because it is known to be robust^{66,67}. We parameterized separate H_t -DBH relationships for four different combinations of edaphic forest type and biogeographical region: (1) terra firme forest in West Africa, (2) terra firme forest in Lower Guinea and the Western Congo Basin, (3) terra firme forest in Eastern Congo Basin and East Africa, (4) seasonally flooded forest from Lower Guinea and the Western Congo Basin (there were no seasonally flooded forest plots in the other biogeographical regions). The parameters are: (1) terra firme forest in West Africa, $a = 56.0$; $b = 0.0401$; $c = 0.744$; (2) terra firme forest in Lower Guinea and the Western Congo Basin, $a = 47.6$; $b = 0.0536$; $c = 0.755$; (3) terra firme forest in the Eastern Congo Basin and East Africa, $a = 50.8$; $b = 0.0499$; $c = 0.706$; and finally (4) seasonally flooded forest from Lower Guinea and the Western Congo Basin, $a = 38.2$; $b = 0.0605$; $c = 0.760$. For each of these combinations of forest type and bioregion, the `local.heights` function combines all height measurements from all plots belonging to that forest type/bioregion and fits the Weibull model parameters using nonlinear least squares (`nls` function in R with default settings), with starting values of $a = 25$, $b = 0.05$ and $c = 0.7$ chosen because they led to regular model convergence. We fitted these models either treating each observation equally or with weights proportional to each tree's basal area. These weights give more importance to large trees during model fitting. We selected the best fitting of these models, determining this to be the model that minimized prediction error of stand biomass when calculated with estimated heights or observed heights. In this way, we selected the non-weighted model for terra firme forests in Lower Guinea/Western Congo Basin and for flooded forests in the Lower Guinea/Western Congo Basin; we selected the weighted model for the other two biogeographical regions (West Africa and Eastern Congo Basin/East Africa). The parameters were used to estimate H_t from DBH for all tree DBH measurements for input into the allometric equation. Median measured individual total tree height is 20.5 m; the height range is 3.1 to 72.5 m. The root mean squared error (RMSE) between the full dataset of measured heights and the predicted heights is 5.7 m, which is 8.0% of the total range. Furthermore, RMSE is 5.3 m in terra firme forest in West Africa (7.5% of the range; $n = 9,771$ trees); RMSE is 6.4 m in terra firme forest in Lower Guinea and the Western Congo Basin (8.7% of the range; $n = 10,838$ trees); RMSE is 4.8 m in terra firme forest in the Eastern Congo Basin and East Africa (8.8% of the range; $n = 3,269$ trees); and RMSE is 4.1 m in seasonally flooded forest from Lower Guinea and the Western Congo Basin (12.5% of the range; $n = 392$ trees).

Wood density. Dry wood density (ρ) measurements were compiled for 730 African species from published sources and stored in www.ForestPlots.net; most were sourced from the Global Wood Density Database on the Dryad digital repository (www.datadryad.org)^{68,69}. Each individual in the tree inventory database was matched to a taxon-specific mean wood density value. Species in both the tree inventory and wood density databases were standardized for orthography and

synonymy using the African Plants Database (www.ville-ge.ch/cjb/bd/africa/) to maximize matches¹³. For incompletely identified individuals or for individuals belonging to species not in the ρ database, we used the mean ρ value for the next-highest known taxonomic category (genus or family, as appropriate). For unidentified individuals, we used the mean wood density value of all individual trees in the plot^{13,52}.

Allometric equation. For each tree we used a published allometric equation⁶¹ to estimate aboveground biomass. We then converted this to carbon, assuming that aboveground carbon (AGC, in Mg C ha⁻¹) is 45.6% of aboveground biomass⁷⁰. Thus:

$$\text{AGC} = 0.456 \times (0.0673 \times (\rho \times \text{DBH}^2 \times H)^{0.976}) / 1,000 \quad (2)$$

with DBH in mm, dry wood density ρ in g cm⁻³, and total tree height H in m (ref. ⁶¹). Aboveground carbon in living biomass for each plot at each census date was estimated as the sum of the AGC of each living stem, divided by plot area (in hectares).

Carbon gain and carbon loss estimation

Net carbon sink (in Mg C ha⁻¹ yr⁻¹) is estimated as carbon gains minus carbon losses. Carbon gains (in Mg C ha⁻¹ yr⁻¹) are the sum of the aboveground live biomass carbon additions from the growth of surviving stems and the addition of newly recruited stems (recruits are stems reaching a DBH ≥ 100 mm during a given census interval), divided by the census length (in years) and plot area (in hectares). For each stem that survived a census interval, carbon additions from its growth (Mg C ha⁻¹ yr⁻¹) were calculated as the difference between its AGC at the end census of the interval and its AGC at the beginning census of the interval. For each stem that recruited during the census interval (that is, reaching DBH ≥ 100 mm), carbon additions were calculated in the same way, assuming DBH = 0 mm at the start of the interval, following standard procedures^{6,65}. Carbon losses (in Mg C ha⁻¹ yr⁻¹) are estimated as the sum of aboveground biomass carbon from all stems that died during a census interval, divided by the census length (in years) and plot area (in hectares). Both carbon gains and carbon losses are calculated using standard methods⁶, including a census interval bias correction, using the SummaryAGWP function of the R package BiomasaFP^{63,64,68}.

As carbon gains (and losses, see below) are affected by a census interval bias, with the underestimate increasing with census length, we corrected this bias by accounting for (1) the carbon additions from trees that grew before they died within an interval (unobserved growth) and (2) the carbon additions from trees that reached 100 mm DBH (that is, were recruited) and then died within the same interval (unobserved recruitment)^{65,71}.

The first component, the unobserved growth of a stem that died during a census interval, is estimated as the difference between AGC at death and AGC at the start of the census. These are calculated using equation (2), from DBH_{death} and DBH_{start}, respectively. The latter is part of the data, the first can be estimated as: $\text{DBH}_{\text{death}} = \text{DBH}_{\text{start}} \times G \times Y_{\text{mean}}$, where G is the plot-level median diameter growth rate (in mm yr⁻¹) of the size class the tree was in at the start of the census interval (size classes are defined as DBH < 200 mm, 200 mm \leq DBH < 400 mm and DBH \geq 400 mm) and Y_{mean} is the mean number of years that trees survived in the census interval before dying. Y_{mean} is calculated from the number of trees that are expected to have died in each year of the census interval, which is derived from the plot-level per capita mortality rate (m_a ; as percentage of dead trees per year) calculated following equation (5) in ref. ⁷¹.

The second component, the growth of recruits that were not observed because they died during the census interval, is estimated by calculating the number of unobserved recruits and diameter at death for each unobserved recruit. The number of unobserved recruits in a given year (stems ha⁻¹ yr⁻¹) is estimated as: $N_{\text{u.r}} = R_a - P_{\text{surv}} \times R_a$, where R_a (number of recruited stems ha⁻¹ yr⁻¹) is the per-area annual recruitment

calculated following equation (11) in ref. ⁷¹ and P_{surv} is the probability of each recruit surviving until the next census: $P_{\text{surv}} = (1 - m_a)^T$, where T is the number of years remaining in the census interval. Summing $N_{\text{u.r}}$ for each year in a census interval gives the total number of unobserved recruits in that census interval. We then estimate diameter at death for each unobserved recruit, which is given in millimetres by $\text{DBH}_{\text{death,u.r}} = 100 + (G_s \times Y_{\text{mean-rec}})$, where G_s is the plot-level median diameter growth rate (in mm yr⁻¹) of the smallest size class (that is, DBH < 200 mm) and $Y_{\text{mean-rec}}$ is the mean number of years that unobserved recruits survived in the census interval before dying. $Y_{\text{mean-rec}}$ is calculated as follows: from m_a we can calculate the number of recruits in a given year that died in each subsequent year, and from this calculate the mean lifespan of recruits in a given year that died before the next census; $Y_{\text{mean-rec}}$ is then the mean of each year's recruit-lifespan, weighted by the number of unobserved recruits in each year.

The census interval bias correction (components one and two combined) typically add <3% to plot-level carbon gains calculated for each plot census interval. Carbon losses are affected by the same census interval bias, so we corrected this bias by accounting for the additional carbon losses from the trees that were recruited and then died within the same interval, and the additional carbon losses resulting from the growth of the trees that died in the interval^{6,15,63}. These two components are calculated in the same way as for carbon gains and typically add <3% to plot-level carbon losses.

Carbon gains include both gains from the growth of surviving stems and new recruits. Separating carbon gains from the tree growth of surviving stems and newly recruited stems shows that carbon gains from recruitment are small overall, and are significantly lower in Africa than in the Amazon (in Africa, 0.17 Mg C ha⁻¹ yr⁻¹; CI: 0.16–0.18 versus in the Amazon, 0.27 Mg C ha⁻¹ yr⁻¹; CI: 0.25–0.28, $P < 0.001$; two-way Wilcoxon test), but this is compensated by carbon gains from survivors being significantly larger in Africa (2.33 Mg C ha⁻¹ yr⁻¹; CI: 2.27–2.39) than in the Amazon (2.13 Mg C ha⁻¹ yr⁻¹; CI: 2.09–2.17, $P = 0.014$). Therefore, gains overall (sum of gains from surviving stems and newly recruited stems) are indistinguishable between the continents (in Africa, 2.57 Mg C ha⁻¹ yr⁻¹; CI: 2.51–2.67 versus in the Amazon, 2.46 Mg C ha⁻¹ yr⁻¹; CI: 2.41–2.50, $P = 0.460$; two-way Wilcoxon test). The lower carbon gains from recruitment in Africa are probably due to the lower stem turnover rates and longer CRT.

Long-term gain, loss and net carbon sink trend estimation

The estimated mean and uncertainty in carbon gains, carbon losses and the net carbon sink of the African plots from 1 January 1983 to 31 December 2014 (Fig. 1, Extended Data Fig. 7 and Extended Data Fig. 8) were calculated following ref. ⁶ to allow direct comparison with published Amazonian results. First, each census interval value was interpolated for each 0.1-year period within the census interval. Then, for each 0.1-year period between 1 January 1983 and 31 December 2014, we calculated a weighted mean of all plots monitored at that time, using the square root of plot area as a weighting factor⁶. Confidence intervals for each 0.1-year period were bootstrapped.

Trends in carbon gains, losses and the net carbon sink over time were assessed using linear mixed effects models (lmer function in R, lme4 package⁷²), providing the linear slopes reported in Fig. 1. These models regress the midpoint of each census interval against the value of the response variable for that census interval. Plot identity was included as a random effect, that is, by assuming that the intercept can vary randomly among plots. We did not include slope as a random effect, consistent with previously published Amazon analyses⁶, because models did not converge owing to some plots having too few census intervals. Observations were weighted by plot size and census interval length. Weighting for the Africa data was derived empirically, by assuming a priori that there is no significant relation between the net carbon sink and census interval length or plot size, following ref. ¹⁵. The following weighting removes all pattern in the residuals:

$$\text{Weight} = \sqrt[3]{\text{length}_{\text{int}}} + \sqrt[4]{\text{plotsize}} - 1 \quad (3)$$

where $\text{length}_{\text{int}}$ is the length of the census interval, in years. Significance was assessed by regressing the residuals of the net carbon sink model against the weights ($P = 0.702$). Similar published weighting was used for the Amazon plots⁶.

Differences in long-term slopes between the two continents for carbon gains, carbon losses and net carbon sink, reported in the main text, were also assessed using linear mixed effects models and weighting, as described above, but performed on the combined African and Amazonian datasets and limited to their common time window, 1 January 1983 to mid-2011. For these three tests on the pooled data (gains, losses and net sink) we included an additional interaction term between census interval date and continent, where a significant interaction would indicate that the slopes differ between continents. The statistical significance of continental differences in slope were assessed using the *F*-statistic (ANOVA function in R, car package⁷³). Shortening the common time window to the 20 years when the continents are best-sampled, mid-1991 to mid-2011, gave very similar results, including a divergent continental sink ($P = 0.04$).

Continental and pan-tropical carbon sink estimates

The per unit area total net carbon sink (in $\text{Mg C ha}^{-1} \text{yr}^{-1}$) for each time period in Table 1 (each decade between 1 January 1980 and 31 December 2009; and between 1 January 2010 and 31 December 2014) is the sum of three components. The first component is the per unit area aboveground carbon sink from living trees and lianas with $\text{DBH} \geq 100$ mm. For Africa we use the per unit area net carbon sink values presented in this paper. For Amazonia, we use data in ref. ⁶. For Southeast Asia, we use inventory data collected using similar standardized methods from 49 plots in ref. ¹⁵. For each time window, we use all plots for which census dates overlap the period, weighted by the square root of plot area, as for the solid lines in Fig. 1. The second component is the per unit area aboveground carbon sink from living trees and lianas with $\text{DBH} < 100$ mm. This is calculated as 5.19%, 9.40% and 5.46% of the first component (that is, aboveground carbon of large living trees) in Africa, Amazonia and Southeast Asia respectively⁷⁴. The third component is the per unit area belowground carbon sink in live biomass, that is, roots. This is calculated as 25%, 37% and 17% of the aboveground carbon of living trees with $\text{DBH} \geq 100$ mm in Africa¹³, Amazonia⁶ and Southeast Asia⁷⁵ respectively.

For each time period in Table 1 we calculated the continental-scale total carbon sink (Pg C yr^{-1}) by multiplying the per unit area total net carbon sink described above by the area of intact forest on each continent at that time interval (in ha) reported in Extended Data Table 2. Decades are calculated from 1 January 1990 to 31 December 1999. For comparability with previous continental-sink results, we used continental values of intact forest area for 1990, 2000, 2005 and 2010 as published in ref. ¹, that is, total forest area minus forest regrowth. We used the 1990–2010 data to fit an exponential model for each continent and used this model to estimate intact forest area for 1980 and 2015.

Finally, in the main text we calculated the proportion of anthropogenic CO_2 emissions removed by Earth's intact tropical forests, as the total pan-tropical carbon sink from Table 1 divided by the total anthropogenic CO_2 emissions. Total anthropogenic CO_2 emissions are calculated as the sum of emissions from fossil fuel and land-use change and are estimated at 7.6 Pg C yr^{-1} in the 1990s, 9.0 Pg C yr^{-1} in the 2000s, and $11.1 \text{ Pg C yr}^{-1}$ in the 2010s (ref. ²¹, assuming 1.7% growth in fossil fuel emissions in 2018 and 2019, and mean 2010–2017 land-use change emissions for 2018 and 2019).

Carbon sink from an atmospheric perspective

To estimate the evolution of the carbon sink from an atmospheric perspective, we assumed that the contribution to the atmosphere

from carbon gains are experienced immediately, while the contribution to the atmosphere from carbon losses must take into account the delay in decomposition of dead trees. We did this by calculating total forest carbon loss ($\text{Mg C ha}^{-1} \text{yr}^{-1}$) for each year in the period 1 January 1950 to 31 December 2014, using the mean 1 January 1983 to 31 December 2014 records from Fig. 1 and assuming constant losses before 1983 ($1.9 \text{ Mg C ha}^{-1} \text{yr}^{-1}$ and $1.5 \text{ Mg C ha}^{-1} \text{yr}^{-1}$ for Africa and Amazonia respectively). Then, for each focal year in the period 1950–2014, we calculated how much carbon was released to the atmosphere in the subsequent years as follows: $y_i = x_0 \times e^{-0.17(i-1)} - x_0 \times e^{-0.17i}$, where x_0 is the total forest carbon loss of the focal year; y_i is the carbon released to the atmosphere at i years from the focal year; and -0.17 yr^{-1} is a constant decomposition rate calculated for tropical forests in the Amazon⁴⁵. For example, carbon loss was $1.95 \text{ Mg C ha}^{-1}$ in 1990 in African forests (Fig. 1), from which $0.31 \text{ Mg C ha}^{-1}$ was released to the atmosphere in 1991; $0.26 \text{ Mg C ha}^{-1}$ in 1992; $0.22 \text{ Mg C ha}^{-1}$ in 1993; $0.07 \text{ Mg C ha}^{-1}$ in 2000 and $0.01 \text{ Mg C ha}^{-1}$ in 2010. Hence, of the full $1.95 \text{ Mg C ha}^{-1}$ dead tree biomass from 1990, $\sim 50\%$ was released to the atmosphere after 4 years, $\sim 85\%$ after 10 years, and $\sim 97\%$ after 20 years. Finally, for each year between 1983 and 2014, the total contribution to the atmosphere from carbon losses was calculated as the sum of all carbon contributions released at that year, including all carbon loss pools from previous years that are released during the focal year (an approach similar to ref. ⁶). We then calculated decadal-scale mean contributions to the atmosphere from carbon losses to estimate the carbon sink from an atmospheric perspective, reported in the main text.

Predictor variable estimates (1983–2015)

For each census interval of each plot, we examined potential predictor variables that may explain the long-term trends in carbon gains and carbon losses, reported in Table 2 and Extended Data Table 1. First, the environmental conditions during the census interval; second, the rate of change of these parameters; and third, forest attributes that may affect how different forests respond to the same environmental change. The predictor variable estimates for each census need to avoid bias due to seasonal variation, for example the intra-annual variability in atmospheric CO_2 concentration. We therefore applied the following procedure to avoid seasonal variability impacts on long-term trends: (1) the length of each focal census interval was rounded to the nearest complete year (for example, a 1.1-year interval became a 1 year interval); (2) we computed dates that minimized the difference between actual fieldwork dates and complete-year census dates, while ensuring that subsequent census intervals of a plot do not overlap. The resulting sequence of non-overlapping census intervals was used to calculate interval-specific means for each environmental predictor variable to remove seasonal effects. The mean difference between the actual fieldwork dates and the complete-year census dates is 0.13 decimal years.

The first group of potential predictor variables, estimated for each census interval of each plot, are theory-driven choices: atmospheric CO_2 concentration, MAT and drought intensity, which we quantified as MCWD^{14,20,76,77}.

Atmospheric CO_2 concentration. CO_2 (in ppm) is estimated as the mean of the monthly mean values from the Mauna Loa record⁷⁸ over the complete year census interval. While atmospheric CO_2 concentration is highly correlated with time ($R^2 = 0.98$), carbon gains are slightly better correlated with CO_2 ($R_{\text{adj}}^2 = 0.0027$) than with time ($R_{\text{adj}}^2 = 0.0025$), as expected from theory.

Mean annual temperature. MAT (in $^{\circ}\text{C}$) was derived from the temporally resolved (1901–2015) dataset of monthly mean temperature from the Climatic Research Unit (CRU TS version 4.03; $\sim 3,025\text{-km}^2$ resolution; released 15 May 2019; <https://crudata.uea.ac.uk/cru/data/hrg/>)⁷⁹. We downsampled the data to $\sim 1\text{-km}^2$ resolution using the WorldClim v2 dataset^{51,80}, by subtracting the difference in mean monthly temperature,

and applying this monthly correction to all months⁸¹. We then calculated MAT for each complete year census interval of each plot using the downscaled monthly CRU record.

Maximum climatological water deficit. MCWD (in mm) was derived from the ~3,025-km² resolution Global Precipitation Climatology Centre dataset (GPCC version 6.0) that includes many more rain gauges than CRU in tropical Africa^{82,83}. Because GPCC ends in 2013 we combined it with satellite-based Tropical Rainfall Measurement Mission data (TRMM 3B43 V7 product, ~757-km² resolution)⁸⁴. The fit for the overlapping time period (1998–2013) was used to correct any systematic difference between GPCC and TRMM: $GPCC' = a + b \times GPCC$, with $GPCC'$ the adjusted GPCC record and a and b being different parameters for each month of the year and for each continent. Precipitation was then downscaled to ~1-km² resolution using the WorldClim dataset^{51,80}, by dividing by the ratio in mean monthly rainfall, and applying this monthly correction to all months⁸¹. For each census interval we extracted monthly precipitation values and estimated evapotranspiration to calculate monthly climatological water deficit (CWD), a commonly used metric of dry season intensity for tropical forests^{14,76,77}. Monthly CWD values were calculated for each subsequent series of 12 months (complete years)⁷⁷. Monthly CWD estimation begins with the wettest month of the first year in the interval, and is calculated as 100 mm per month evapotranspiration (ET) minus monthly precipitation (P). Then, CWD_{*i*} values for the subsequent 11 months (i) were calculated recursively as: $CWD_i = ET - P_i + CWD_{i-1}$, where negative CWD_{*i*} values were set to zero⁷⁷ (no drought conditions). This procedure was repeated for each subsequent complete 12 months. We then calculated the annual MCWD as the largest monthly CWD value for every complete year within the census interval, with the MCWD of a census interval being the mean of the annual MCWD values within the census interval. Larger MCWD indicates more severe water deficits.

We assume evapotranspiration is 100 mm per month on both continents, based on measurements from Amazonia^{76,77}, more limited measurements from West Africa summarized in ref.⁸⁵, predictive skill⁸⁶, and use in past studies on both continents^{14,87}. MCWD therefore represents a precipitation-driven dry season deficit, given that evapotranspiration remains constant. An alternative assessment, using a data-driven evapotranspiration product^{88,89}, gave a mean evapotranspiration of 95 mm and 98 mm per month for the African and Amazonian plot networks respectively (mean for the 1982–2008 period). Using these values did not affect the results.

To calculate the environmental change of potential predictor variables, CO₂-change (in ppm yr⁻¹), MAT-change (in °C yr⁻¹) and MCWD-change (in mm yr⁻¹), we selected an optimum period over which to calculate the change, derived empirically by assessing the correlation of carbon gains (all plots, all censuses) with the change in each environmental variable, using linear mixed effects models (lmer function in R, lme4 package⁷²). The annualized change in the environmental variable was calculated as the change between the focal interval and a prior interval (termed the baseline period) with a lengthening time window ranging from 1 year through to 80 years before the focal interval (that is, 80 linear mixed effects models per variable). We calculated Akaike's Information Criterion (AIC) for each model and selected the interval length with the lowest AIC. Thus, $MAT\text{-}change = (MAT_i - MAT_b) / (date_i - date_b)$, where MAT_i is the MAT over the focal census interval calculated using the procedure described above, MAT_b is the MAT over a baseline period before the focal interval, $date_i$ is the mid-date of the focal census interval and $date_b$ is the mid-date of the baseline period. The lmer results show that the baseline period for MAT-change is 5 years and for CO₂-change it is 56 years, while MCWD showed no clear trend, so MCWD-change was not included in the models (see Extended Data Fig. 3). All three results conform to a priori theoretical expectations. For CO₂ a maximum response to an integrated 56 years of change is expected because forest stands will respond most strongly to CO₂ when

most individuals have grown under the new rapidly changing condition, which should be at its maximum at a time approximately equivalent to the CRT of a forest stand^{30,90} (mean of 62 years in the pooled dataset). For MAT, 5 years is consistent with experiments showing temperature acclimation of leaf- and plant-level photosynthetic and respiration processes over half-decadal timescales^{31,91}. MCWD has no overall trend suggesting that once a drought ends, its impact on tree growth fades rapidly, as seen in other studies^{14,92}. Furthermore, in the moist tropics wet-season rainfall is expected to recharge soil water, so lagged impacts of droughts are not expected.

We calculated estimates of two forest attributes that may alter responses to environmental change as potential predictor variables: wood density and CRT. In intact old-growth forests, mean wood density (in g cm⁻³) is inversely related to resource availability^{28,93,94}, as is seen in our dataset (carbon gains and plot-level mean wood density are negatively correlated; Extended Data Fig. 4). Wood density is calculated for each census interval in the dataset, as the mean wood density of all trees alive at the end of the census interval, to be consistent with the previous Amazon analysis⁶. Carbon residence time (CRT, in years) is a measure of the time that fixed carbon stays in the system. CRT is a potential correlate of the impact of past carbon gains on later carbon losses³⁰. To avoid circularity in the models, the equation used to calculate CRT differed depending on the response variable. If the response variable is carbon loss, the CRT equation is based on gains: $CRT = AGC / \text{gains}$, with AGC for each interval based on AGC at the end of the interval, and the gains for each interval calculated as the time-weighted mean of the gains in the interval and the previous intervals (that is, long-term gains). If the response variable is carbon gains, the CRT equation is based on losses: $CRT = AGC / \text{losses}$. The equation employed for use in the carbon loss model (based on gains) is the standard formula used to calculate CRT and is retained in the minimum adequate model (see below and Table 2). The non-standard CRT equation (based on losses) used in the carbon gain model is not retained in the minimum adequate model (see below).

Statistical modelling of the carbon gain, loss and sink trends

We first constructed two models including those environmental drivers exhibiting long-term change that impact theory-driven models of photosynthesis and respiration as predictor variables: CO₂, MAT and MCWD. One model had carbon gains as the response variable, the other had carbon losses as the response variable (both in Mg C ha⁻¹ yr⁻¹). Models were fitted using the lme function in R, with maximum likelihood (NLME package⁹⁵). All census intervals within all plots were used, weighted by plot size and census length (using equation (3)). Plot identity was included as a random effect, that is, assuming that the intercept can vary randomly among plots. All predictor variables in the models were scaled without centring (scale function in R, RASTER package⁶²). Carbon gain values were normally distributed but carbon loss values required a power-law transformation ($\lambda = 0.361$) to meet normality criteria. Multi-parameter models are: carbon gains = $\text{intcp} + a \times CO_2 + b \times MAT + c \times MCWD$ (model 1); carbon losses = $\text{intcp} + a \times CO_2 + b \times MAT + c \times MCWD$ (model 2); where intcp is the estimated model intercept, and a , b and c are model parameters giving the slope of relationships with environmental predictor variables. For multi-parameter model outputs see Extended Data Table 1, for single-parameter relationships, Fig. 2.

The second pair of models include the same environmental predictors (CO₂, MAT, MCWD), plus their rate of change (CO₂-change, MAT-change, but not MCWD-change, as explained above), and forest attributes that may alter how forests respond to the same environmental change (wood density, CRT), as described above. We also evaluated the possible inclusion of a differential continent effect of each variable in the full model. We first constructed models with only a single predictor variable, and allowed different slopes in each continent. Next, if removal of the continent-specific slope (using stepAIC function in R, MASS package⁹⁶) increased model AIC then the continent-specific

slope was included in the full model for that variable. Only MCWD showed a significant differential continent-specific slope ($P < 0.001$). This implies that forests on both continents have common responses to CO₂, CO₂-change, MAT, MAT-change, wood density and CRT, but respond differently to differences in MCWD. This may be because wet-adapted species are much rarer in Africa than in Amazonia as a result of large differences in past climate variation³⁴. Last, we allowed different intercepts for the two continents to potentially account for differing biogeographical or other continent-specific factors. For the carbon loss model, we applied the same continent-specific effects for slope as for the carbon gain model. Carbon loss values were transformed using a power-law transformation ($\lambda = 0.361$) to meet normality criteria.

For both carbon gains and losses we parameterized a global model including the significant continent-specific effect of MCWD, selecting the most parsimonious simplified model using all-subsets regression^{97,98}. To do so, we first generated a set of models with all possible combinations (subsets) of fixed effect terms in the global model using the dredge function of the MuMIn package in R⁹⁹. We then chose the best-ranked simplified model based on the second-order Akaike Information Criterion (known as AICc), hereafter called the 'minimum adequate carbon gain/loss model', reported in Table 2. The minimum adequate models are: carbon gains = $\text{intcp} \times \text{continent} + a \times \text{CO}_2\text{-change} + b \times \text{MAT} + c \times \text{MAT-change} + d \times \text{MCWD} \times \text{continent} + e \times \text{wood density}$ (model 3); carbon losses = $\text{intcp} + a \times \text{CO}_2\text{-change} + b \times \text{MAT-change} + c \times \text{MCWD} + d \times \text{CRT}$ (model 4). Wood density was retained in the carbon gain model, probably because growth is primarily affected by resource availability, whereas CRT was retained in the carbon loss model, probably because losses are primarily affected by how long fixed carbon is retained in the system.

Table 2 presents model coefficients of the best-ranked gain model and best-ranked loss model selected using all-subsets regression. These best-ranked gain and loss models have weights of 0.310 and 0.132 respectively, which is almost double the weight of the second-ranked models (0.152 and 0.075 respectively). In Supplementary Table 2 we also used the model.avg function of the MuMIn package to calculate a weighted mean of the coefficients of the models that together represent a cumulative weight-sum of 0.95 (that is, a 95% confidence subset). Supplementary Table 2 (model-averaged) and Table 2 (best-ranked) model parameters are very similar. Supplementary Tables 3 and 4 report the complete sets of carbon gains and loss models that contribute to the model average results.

The model-average results show the same continental differences in sensitivity to environmental variables as the best-ranked models. From 1 January 2000 to 31 December 2014, carbon gains increased owing to CO₂-change (+3.7% in both the averaged and the best-ranked models, both continents), whereas temperature rises led to a decline in gains, which especially had an effect in the Amazon (−1.14% and −1.07% due to MAT and MAT-change together in the averaged and best-ranked model respectively). Finally, both model-average and best-ranked models result in similar predictions of the net carbon sink over the 1 January 1983 to 31 December 2039 period: the future net sink trend in Africa is −0.004 and −0.003 in the best-ranked and averaged models, respectively; in Amazonia the future net sink trend is −0.013 and −0.011 in the best-ranked and averaged models, respectively. The Amazon sink reaches zero in 2041 using model-averaged parameters compared to 2035 using the best-ranked models.

Estimating future predictor variables to 2040

To calculate future modelled trends in carbon gains and losses (Fig. 3), we first estimated annual records of the predictor variables (CO₂-change, MAT, MAT-change, MCWD, wood density and CRT) to 31 December 2039 (Extended Data Fig. 5).

To do so, we first calculated annual records for the period of the observed trends for each plot location (that is, from 1 January 1983 to 31 December 2014 in Africa and 1 January 1983 to mid-2011 in Amazonia).

For CO₂-change, MAT, MAT-change and MCWD we extracted monthly records as described in the Methods section 'Predictor variable estimates (1983–2014)'. For wood density and CRT we interpolated to a 0.1-year period within each census interval (as in Fig. 1). Then, we calculated the mean annual value of each predictor variable from the 244 plot locations in Africa, and separately the mean annual value of each predictor variable from the 321 plot locations in Amazonia (solid lines in Extended Data Fig. 5). For each predictor variable, we calculated annual records of upper and lower confidence intervals by respectively adding and subtracting 2σ to the mean of each annual value (shaded area in Extended Data Fig. 5). Second, for each predictor variable we parameterized a linear model for each continent using the annual records for the period of the observed trends. Then for each predictor variable, the continent-specific linear regression models were used to estimate predictor variables for each plot location from 1 January 2015 to 31 December 2039 in Africa and from mid-2011 to 31 December 2039 in the Amazon (dotted lines in Extended Data Fig. 5).

Estimating future carbon gain, loss and sink trends

We used the minimum adequate models (Table 2) to predict annual records of carbon gain, carbon loss and the carbon sink for the plot networks in Africa and Amazonia over the period 1983 through to 2040 (Fig. 3). We extracted predicted carbon gain and loss values using the mean annual records for each predictor variable (predictSE.lme function, AICcmodavg package¹⁰⁰). Upper and lower confidence intervals were calculated accounting for uncertainties in the model (both fixed and random effects) and predictor variables using the 2σ upper and lower confidence interval for each predictor variable (using predictSE.lme). Finally, the net carbon sink was calculated by subtracting the losses from the gains. To obtain sink values in the future, reported in Table 1, annual per unit area sink predictions (from Fig. 3) were averaged over each decade and multiplied by the future forest area, as described above.

To test the sensitivity of the future predictions in Fig. 3, we reran the analysis by modifying future trajectories of predictor variables one at a time, while keeping all others the same, to assess the mean C sink over 2010–15 and 2030 (averaging at 2030 is not necessary as trends in MAT-change and MCWD, which largely drive modelled inter-annual variability, are estimated as smooth trends in the future). For each predictor variable, we explored the potential impacts of the likely bounds of possibility: (1) by taking the steepest slope of either continent from the extrapolated trends, doubling this slope and applying it on both continents; and (2) by taking the steepest slope of either continent from the extrapolated trends, taking the additive inverse of this slope and applying it on both continents. These bounds represent deviations of $>2\sigma$ from observed trends. Change in MAT also alters MAT-change, so we present the sensitivity of both parameters together.

Additionally, for CO₂-change and MAT, we also calculated future slopes under three future Representative Concentration Pathway (RCP) scenarios³⁸ with different radiative forcing in 2100: RCP2.6, RCP4.5 and RCP8.5. Future RCP CO₂-change slopes (ppm yr^{−1}) were calculated using RCP CO₂ concentration data for the years between 2015 and 2030 inclusive. Future RCP MAT and MAT-change slopes were obtained from plot-specific MAT values extracted from downscaled 1-km² resolution data for current⁸⁰ and future⁵¹ climate from WorldClim, and averaged over 19 CMIP5 models. We subtracted the mean 2040–2060 climate MAT (that is, 2050) from the mean 1970–2000 climate MAT (that is, 1985), divided by 65 years to give the annual rate of change. We then calculated a mean slope over all plots per continent. Finally, to avoid mismatches between RCP-derived values of CO₂ and MAT and the observed records, we removed any difference in intercept between the RCP trends and observed trends, so that the RCP trends were a continuation of the end-point of the observed trajectory (31 December 2014). We did not estimate the sensitivity of MCWD under the RCP scenarios, because the mean of the CMIP5 models do not show drought trends for our forest

plot networks, unlike rain gauge data for the recent past^{41,42}, and thus would show little or no sensitivity to MCWD. For each modified slope, Supplementary Table 5 reports the absolute decline in the sink in each continent in 2030 compared to the 2010–15 mean sink. This shows that the future sink strength is sensitive to future environmental conditions, but within both RCP scenarios and our bounds of possibility we show a decline in the sink strength in both continents over the 2020s.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source data to generate figures and tables are available from https://doi.org/10.5521/Forestplots.net/2019_1.

Code availability

R code to generate figures and tables is available from: https://doi.org/10.5521/Forestplots.net/2019_1.

51. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
52. Lopez-Gonzalez, G., Lewis, S. L., Burkitt, M. & Phillips, O. L. ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. *J. Veg. Sci.* **22**, 610–613 (2011).
53. Lopez-Gonzalez, G., Lewis, S. L., Burkitt, M., Baker, T. R. & Phillips, O. L. *ForestPlots.net Database* <http://www.forestplots.net> (2009).
54. Sheil, D. & Bitariho, R. *Bwindi Impenetrable Forest TEAM Site* <https://www.wildlifesights.org/team-network>, TEAM-DataPackage-20151201235855_1254 (2009).
55. Kenfack, D. *Korup National Park TEAM Site* <https://www.wildlifesights.org/team-network>, TEAM-DataPackage-20151201235855_1254 (2011).
56. Rovero, F., Marshall, A. & Martin, E. *Udzungwa TEAM Site* <https://www.wildlifesights.org/team-network>, TEAM-DataPackage-20151130235007_5069 (2009).
57. Hockemba, M. B. N. *Nouabalé Ndoki TEAM Site* <https://www.wildlifesights.org/team-network>, TEAM-DataPackage-20151201235855_1254 (2010).
58. Anderson-Teixeira, K. J. et al. CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Change Biol.* **21**, 528–549 (2015).
59. Gourlet-Fleury, S. et al. Tropical forest recovery from logging: a 24 year silvicultural experiment from Central Africa. *Phil. Trans. R. Soc. Lond. B* **368**, 20120302 (2013).
60. Claeyes, F. et al. Climate change would lead to a sharp acceleration of Central African forests dynamics by the end of the century. *Environ. Res. Lett.* **14**, 044002 (2019).
61. Chave, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Change Biol.* **20**, 3177–3190 (2014).
62. R Development Core Team R: *A Language and Environment for Statistical Computing* <http://www.R-project.org/> (2015).
63. Lopez-Gonzalez, G., Sullivan, M. & Baker, T. BiomasaFP. R package version 0.2.1 <http://www.forestplots.net/en/resources/analysis> (2017).
64. Phillips, O., Baker, T., Brien, R. & Feldpausch, T. RAINFOR field manual for plot establishment and remeasurement. http://www.rainfor.org/upload/ManualsEnglish/RAINFOR_field_manual_version_2016.pdf (Univ. Leeds, 2016).
65. Talbot, J. et al. Methods to estimate aboveground wood productivity from long-term forest inventory plots. *For. Ecol. Manage.* **320**, 30–38 (2014).
66. Sullivan, M. J. P. et al. Field methods for sampling tree height for tropical forest biomass estimation. *Methods Ecol. Evol.* **9**, 1179–1189 (2018).
67. Feldpausch, T. R. et al. Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* **9**, 3381–3403 (2012).
68. Chave, J. et al. Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12**, 351–366 (2009).
69. Zanne, A. E. et al. *Towards a Worldwide Wood Economics Spectrum* <https://doi.org/10.5061/dryad.234> (Dryad Digital Repository, 2009).
70. Martin, A. R., Doraisami, M. & Thomas, S. C. Global patterns in wood carbon concentration across the world's trees and forests. *Nat. Geosci.* **11**, 915–920 (2018).
71. Kohyama, T. S., Kohyama, T. I., Sheil, D. & Rees, M. Definition and estimation of vital rates from repeated censuses: choices, comparisons and bias corrections focusing on trees. *Methods Ecol. Evol.* **9**, 809–821 (2018).
72. Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: linear mixed-effects models using Eigen and S4. R package version 1.0-4 <http://www.inside-r.org/packages/lme4/> versions/1-0-4 (2013).
73. Fox, J. *Applied Regression Analysis and Generalized Linear Models* 2nd edn (Sage Publishing, 2008).
74. Chave, J. et al. Assessing evidence for a pervasive alteration in tropical tree communities. *PLoS Biol.* **6**, 0455–0462 (2008).
75. Yuen, J. Q., Ziegler, A. D., Webb, E. L. & Ryan, C. M. Uncertainty in below-ground carbon biomass for major land covers in Southeast Asia. *For. Ecol. Manage.* **310**, 915–926 (2013).
76. Aragão, L. E. O. C. et al. Spatial patterns and fire response of recent Amazonian droughts. *Geophys. Res. Lett.* **34**, L07701 (2007).
77. Aragão, L. E. O. C. et al. Environmental change and the carbon balance of Amazonian forests. *Biol. Rev. Camb. Phil. Soc.* **89**, 913–931 (2014).
78. Tans, P. & Keeling, R. *Trends in Atmospheric Carbon Dioxide for Mauna Loa, Hawaii* <http://www.esrl.noaa.gov/gmd/ccgg/trends/> (ESRL, 2016).
79. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
80. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
81. Ramirez-Villegas, J. & Jarvis, A. *Downscaling Global Circulation Model Outputs: The Delta Method*. Decision and Policy Analysis Working Paper No. 1 <https://cgspage.cgiar.org/handle/10568/90731> (International Center for Tropical Agriculture (CIAT), 2010).
82. Schneider, U. et al. *GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data* https://opendata.dwd.de/climate_environment/GPCC/html/fulldata_v6_doi_download.html (Global Precipitation Climatology Centre (GPCC) at Deutscher Wetterdienst, 2011).
83. Sun, Q. et al. Review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev. Geophys.* **56**, 79–107 (2017).
84. Huffman, G. J. et al. The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**, 38–55 (2007).
85. Kume, T. et al. Ten-year evapotranspiration estimates in a Bornean tropical rainforest. *Agric. For. Meteorol.* **151**, 1183–1192 (2011).
86. Zelazowski, P., Malhi, Y., Huntingford, C., Sitch, S. & Fisher, J. B. Changes in the potential distribution of humid tropical forests on a warmer planet. *Phil. Trans. R. Soc. A* **369**, 137–160 (2011).
87. James, R., Washington, R. & Rowell, D. P. Implications of global warming for the climate of African rainforests. *Phil. Trans. R. Soc. Lond. B* **368**, 20120298 (2013).
88. Jung, M. et al. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **467**, 951–954 (2010).
89. Jung, M. et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.* **116**, <https://doi.org/10.1029/2010JG001566> (2011).
90. Lloyd, J. & Farquhar, G. D. The CO₂ dependence of photosynthesis, plant growth responses to elevated atmospheric CO₂ concentrations and their interaction with soil nutrient status. I. General principles and forest ecosystems. *Funct. Ecol.* **10**, 4–32 (1996).
91. Aspinwall, M. J. et al. Convergent acclimation of leaf photosynthesis and respiration to prevailing ambient temperatures under current and warmer climates in *Eucalyptus tereticornis*. *New Phytol.* **212**, 354–367 (2016).
92. Bonal, D., Burbanck, B., Stahl, C., Wagner, F. & Hérault, B. The response of tropical rainforests to drought—lessons from recent research and future prospects. *Ann. For. Sci.* **73**, 27–44 (2016).
93. Quesada, C. A. et al. Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences* **7**, 1515–1541 (2010).
94. Baker, T. R., Swaine, M. D. & Burslem, D. F. R. P. Variation in tropical forest growth rates: combined effects of functional group composition and resource availability. *Perspect. Plant Ecol. Evol. Syst.* **6**, 21–36 (2003).
95. Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* 1st edn 528 (Springer, 2000).
96. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* 4th edn 498 (Springer, 2002).
97. Olejnik, S., Mills, J. & Keselman, H. Using Wherry's adjusted R² and Mallows' Cp for model selection from all possible regressions. *J. Exp. Educ.* **68**, 365–380 (2000).
98. Whittingham, M. J., Stephens, P. A., Bradbury, R. B. & Freckleton, R. P. Why do we still use stepwise modelling in ecology and behaviour? *J. Anim. Ecol.* **75**, 1182–1189 (2006).
99. Bartoň, K. MuMin: Multi-Model Inference. Tools for performing model selection and model averaging. R package version 1.43.6 (2019).
100. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, 2007).
101. Mayaux, P., De Grandi, G. & Malingreau, J.-P. Central African forest cover revisited: a multisatellite analysis. *Remote Sens. Environ.* **71**, 183–196 (2000).
102. Mayaux, P. et al. *The Land Cover Map for Africa in the Year 2000* GLC2000 database, <https://forobs.jrc.ec.europa.eu/products/glc2000/products.php> (European Commission Joint Research Centre, 2003).

Acknowledgements This paper is a product of the African Tropical Rainforest Observatory Network (AfriTRON), curated at ForestPlots.net. AfriTRON has been supported by numerous people and grants since its inception. We sincerely thank the people of the many villages and local communities who welcomed our field teams and without whose support this work would not have been possible: Sierra Leone (villages: Barrie, Gaura, Koya, Makpele, Malema, Nomo, Tunkia; teams in protected areas: the Gola Rainforest National Park), Liberia (villages: Garley town, River Gbeh, Glaro Freetown), Ghana (villages: Nkwanta, Asenanyo, Bonsa, Agona, Boekrom, Dadieso, Enchi, Dabiasem, Mangowase, Draw, Fure, Esuboni, Okumaninin, Kade, Asamankese, Tinte Bepo, Tonton), Nigeria (Oban village), Gabon (villages: Ekobakoba, Mikongo, Babilone, Makokou, Tchimbelle, Mondah, Ivindo, Ebe, Ekouk, Oveng, Sette Cama; teams in protected areas: Ivindo National Park, Loke National Park, Waka National Park; teams in concessions: Ipasa station, Kingele station, Leke/Moyabi Rougier Forestry Concession), Cameroon (villages: Campo, Nazareth, Lomié, Djomédjo, Alat-Makay, Somalomo, Deng Deng, Eyumojok, Mbakaou, Myere, Nguti, Bejange, Kepkane, Basho, Mendhi, Matene, Mboh, Takamanda, Obonyi, Ngoila; teams in protected areas: Ejagham forest reserve), Democratic Republic of Congo (villages: Yoko, Yangambi, Epulu, Monkoto), Republic of Congo (villages: Bomassa, Ekolongouma, Bolembé, Makao, Mbéli, Kabo, Niangué, Ngubu, Goulaki, Essimbi). We thank the field assistants whose expertise and enthusiasm is indispensable to successful fieldwork, including: M. E. Abang, U. P. Achui, F. Addai, E. J. Agbachon, J. Agnaka, A. J. Akaza, G. Alaman, G. Alaman, A. E. Alexander, K. Allen, M. Amalphi, D. Amandus, J. Andju, L. A. Limbanga, S. Asamoah, T. M. Ashu, M. Ashu, J. Asse, B. Augustine, H. Badjoko, M. Balimu, J. Baviogui-Baviogui, S. Bente, A. Bertrand, A. Bettus, A. Bias, A. Bikoula, A. Bimba,

Article

P. Bissiemiou, M. Boateng, E. Bonyenga, M. B. Ekaya, G. Bouka, J. Boussengui, D. B. Ngomo, C. Chalange, S. Chenikan, J. Dabo, E. Dadize, T. Degraft, J. Dibakou, J.-T. Dikangadissi, P. Dimbonda, E. Dimoto, C. Ditougou, D. Dorbor, M. Dorbor, V. Droissart, K. Duah, E. Ebe, O. J. Eji, E. B. Ekamam, J.-R. Ekomindong, E. J. Enow, H. Entombo, E. M. Ernest, C. Esola, J. Essouma, A. Gabriel, N. Genesis, B. Gideon, A. Godwin, E. Grear, D. J. Grear, M. Ismael, M. Iwango, M. Iyafa, N. Kamdem, B. Kibinda, A. Kidimbu, E. Kimumbu, J. Kintsieri, C. K. Opepa, A. Kitegile, T. Komo, P. Koué, A. Kouanga, J. J. Koumikaka, I. Liengola, E. Litonga, L. Louvouando, O. Luis, N. M. Mady, F. Mahoula, A. Mahundu, C. A. Mandebet, P. Maurice, K. Y. Mayossa, R. M. Nkogue, I. D. Mbe, C. Mbina, H. Mbona, A. Mboni, A. Mbouni, P. Menzo, M. Menge, A. Michael, A. Mindoumou, J. Minpsa, J. P. Mondjo, E. Mounoumoulossi, S. Mpouam, T. Msigala, J. Msirikale, S. Mtoka, R. Mwakisoma, D. Ndong-Nguema, G. Ndayame, G. Ngongbo, F. Ngowa, D. Nguema, L. Nguye, R. Niangadouma, Y. Nkrumah, S. Nshimba, M. N. Mboumba, F. N. Obiang, L. Obi, R. Obi, E. L. Odjong, F. Okon, F. Oliveira, A. L. Owemicho, L. Oyen-Amoni, A. Platini, P. Ploton, S. Quausah, E. Ramazani, B. S. Jean, L. Sagang, R. Salter, A. Seki, D. Shirima, M. Simo, I. Singono, A. E. Tabi, T. G. Tako, N. G. Tambe, T. Tcho, A. Teah, V. Tehtoe, B. J. Telephas, M. L. Tonda, A. Tresor, H. Umenendo, R. Votere, C. K. Weah, S. Weah, B. Wursten, E. Yalley, D. Zebaze, L. Cerbonney, E. Dubiez, H. Moinecourt, F. Lanckriet, S. Samai, M. Swaray, P. Lamboi, M. Sullay, D. Bannah, I. Kanneh, M. Kannah, A. Kemokai, J. Kenneh and M. Lukulay. For logistical and administrative support, we are indebted to international, national and local institutions: the Forestry Department of the Government of Sierra Leone, the Conservation Society of Sierra Leone, the Royal Society for the Protection of Birds (RSPB, UK), The Gola Rainforest National Park (Sierra Leone), the Forestry Development Authority of the Government of Liberia (FDA), the University of Liberia, the Forestry Commission of Ghana (FC), the Forestry Research Institute of Ghana (FORIG), University of Ibadan (Nigeria), the University of Abeokuta (Nigeria), the Ministère des Eaux, Forêts, Chasse et Pêche (MEFCP, Central African Republic), the Institut Centrafricain de Recherche Agronomique (ICRA, Central African Republic), The Service de Coopération et d'Actions Culturelles (SCAC/MAE, Central African Republic), The University of Bangui (Central African Republic), the Société Centrafricaine de Déroulage (SCAD, Central African Republic), the University of Yaounde I (Cameroon), the National Herbarium of Yaounde (Cameroon), the University of Buea (Cameroon), Bioversity International (Cameroon), the Ministry of Forests, Seas, Environment and Climate (Gabon), the Agence Nationale des Parcs Nationaux de Gabon (ANPN), Institut de Recherche en Écologie Tropicale du Gabon, Rougier-Gabon, the Marien Ngouabi University of Brazzaville (Republic of Congo), the Ministère des Eaux et Forêts (Republic of Congo), the Ministère de la Recherche Scientifique et de l'Innovation Technologique (Republic of Congo), the Nouabalé-Ndoki Foundation (Republic of Congo), WCS-Congo, Salonga National Park (Democratic Republic of Congo), The Centre de Formation et de Recherche en Conservation Forestière (CEFRECOP, Epulu, Democratic Republic of Congo), the Institut National pour l'Étude et la Recherche Agronomiques (INERA, Democratic Republic of Congo), the École Régionale Postuniversitaire d'Aménagement et de Gestion intégrés des Forêts et Territoires tropicaux (ERAIFT Kinshasa, Democratic Republic of Congo), WWF-Democratic Republic of Congo, WCS-Democratic Republic of Congo, the Université de Kisangani (Democratic Republic of Congo), Université Officielle de Bukavu (Democratic Republic of Congo), Université de Mbuji-Mayi (Democratic Republic of Congo), le Ministère de l'Environnement et Développement Durable (Democratic Republic of Congo), the FORETS project in Yangambi (CIFOR, CGIAR and the European Union; Democratic Republic of Congo), the Lukuru Wildlife Research Foundation (Democratic Republic of Congo), Mbarara University of Science and Technology (MUST, Uganda), WCS-Uganda, the Uganda Forest Department, the Commission of Central African Forests (COMIFAC), the Udzungwa Ecological Monitoring Centre (Tanzania) and the Sokoine University of Agriculture (Tanzania). We thank C. Chatelain (Geneva Botanic Gardens) for access to the African Plants Database. Grants that have funded the AfriTRON network including data in this paper are: a European Research Council Advanced Grant to O.L.P. and S.L.L. (T-FORCES; 291585; Tropical Forests in the Changing Earth System), a NERC grant to O.L.P., Y.M., and S.L.L. (NER/A/S/2000/01002), a Royal Society University Research Fellowship to S.L.L., a NERC New Investigators Grant to S.L.L., a Philip Leverhulme Award to S.L.L., a European Union FP7 grant to E.G. and S.L.L. (GEOCARBON; 283080), Valuing the Arc Leverhulme Program Grant to Andrew Balmford and S.L.L., a Natural Environment Research Council (NERC) Consortium Grant to Jon Lloyd and S.L.L. (TROBIT; NE/DO05590/), the Gordon and Betty Moore Foundation to L.J.T.W. and S.L.L., the David and Lucile Packard Foundation to L.J.T.W. and S.L.L., the Centre for International Forestry Research to T.S. and

S.L.L. (CIFOR), and Gabon's National Parks Agency (ANPN) to S.L.L. W.H. was funded by T-FORCES and the Brain programme of the Belgian Federal Government (BR/132/A1/AFRIFORD grant to Olivier Hardy and the BR/143/A3/HERBAXYLAREDD grant to H.B.). O.L.P., S.L.L., M.J.P.S., A.E.-M., A.L., G.L.-G., G.P. and L.Q. were supported by T-FORCES. Eight plots (codes ANK, IVI, LPG, MNG) included in AfriTRON are also part of the Global Ecosystem Monitoring network (GEM). Additional African data were included from the consortium MEFCP-ICRA-CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement), the Tropical Ecology Assessment and Monitoring Network (TEAM), and the Forest Global Earth Observatory Network (ForestGEO; formerly the Center for Tropical Forest Science, CTFS). The TEAM network is a collaboration between Conservation International, the Missouri Botanical Garden, the Smithsonian Institution and the Wildlife Conservation Society, and funded by the Gordon and Betty Moore Foundation and other donors. The ForestGEO Network is a collaboration between the Smithsonian Institution, other federal agencies of the United States, the Wildlife Conservation Society (WCS) and the World Wide Fund for Nature (WWF), and funded by the US National Science Foundation and other donors. The paper was made possible by the RAINFOR network in Amazonia, with multiple funding agencies and hundreds of investigators working in Amazonia, acknowledged in ref. ⁶, providing comprehensive published data and code and assisting in the onward analysis of their data; see ref. ⁶. Data from AfriTRON and RAINFOR are stored and curated by ForestPlots.net, a long-term cyber-infrastructure initiative hosted at the University of Leeds that unites permanent plot records and their contributing scientists from the world's tropical forests. The development of ForestPlots.net and curation of most data analysed here was funded by many sources, including grants to O.L.P. (principally from ERC AdG 291585 'T-FORCES', NERC NE/B503384/1 and the Gordon and Betty Moore Foundation 1656 'RAINFOR'), T.R.B. (the University of Leeds contribution to 'AMAZALERT', NERC (NE/I028122/1) with T. Pennington, the Gordon and Betty Moore Foundation ('MonANPeru') and a NERC Impact Accelerator grant for the initial development of the BiomasaFPR package), E.G. ('GEOCARBON' and NE/FO05806/1 'AMAZONICA') and S.L.L. (Royal Society University Research Fellowship, NERC New Investigators Award, NERC NE/PO08755/1). We acknowledge the contributions of the ForestPlots.net developers (M. Burkitt, G. Lopez-Gonzalez) and the steering committee (T.R.B., A.L., S.L.L., O.L.P., L.Q., E. N. H. Coronado and B. S. Marimon) for advice on database development and management.

Author Contributions S.L.L. conceived and managed the AfriTRON forest plot recensus programme, O.L.P., T.C.H.S., L.J.T.W. and Y.M. contributed to its development. W.H., S.L.L., O.L.P., B.S. and M.J.P.S. developed the study. W.H., S.L.L., O.L.P., K.A.-B., H.B., A.C.-S., C.E.N.E., S.F., D.S., B.S., T.C.H.S., S.C.T., K.A.A., S.A.-B., C.A.A., T.R.B., L.F.B., F. Baya, S.K.B., F. Benedet, R.B., Y.E.B., P. Boeckx, P. Boundja, T.B., E.C., G.B.C., C.J.C., M.C., J.A.C., D.C., A.K.D., G.C.D., T.d.H., M.D.K., J.-L.D., T.R.F., A.F., E.G.F., M.G., C.G., S.G.-F., J.S.H., A.C.H., D.J.H., T.B.H., M.B.N.H., A.H., S.A.I., K.J.J., T.J., E.K.Y., E.K., D.K., M.E.L., J.A.L., J.L., J.C.L., J.-R.M., Y.M., A.R.M., J.M., E.H.M., F.M.M., V.P.M., V.M., E.T.A.M., S.M., J.M.M., P.K.T.M., N.N.B., L.O., F.E., K.S.-H.P., A.D.P., J.R.P., L.Q., J.R., F.R., M.D.S., H.T., J. Talbot, J. Taplin, D.M.T., D.W.T., B.T., J.T.M., D.T., P.M.U., G.v.d.H., H.V., J.V., L.J.T.W., S.W., H.W., J.T.W. and L.Z. contributed data (with larger field contributions by S.L.L., W.H., A.C.-S., B.S., H.T., A.K.D., C.E.N.E., J.M.M., K.A.-B. and S.F.). O.L.P., T.R.B., S.L.L. and G.L.-G. conceived and managed forestplots.net; O.L.P., T.R.B., S.L.L., E.G., G.L.-G., G.C.P., A.L., R.J.W.B., T.R.F. and M.J.P.S. developed it. W.H., M.J.P.S., S.L.L., O.L.P., R.J.W.B., A.L., G.L.-G., A.E.-M., A.K., E.G., T.R.B., A.C.B. and G.C.P. contributed analysis tools. W.H. and S.L.L. analysed the data (with important contributions from M.J.P.S.). S.L.L. and W.H. wrote the paper. All co-authors read and approved the manuscript (with important insights provided by O.L.P., S.F., R.J.W.B., E.G., H.B., D.S., M.J.P.S., S.G.-F., P.B., H.V. and S.C.T.).

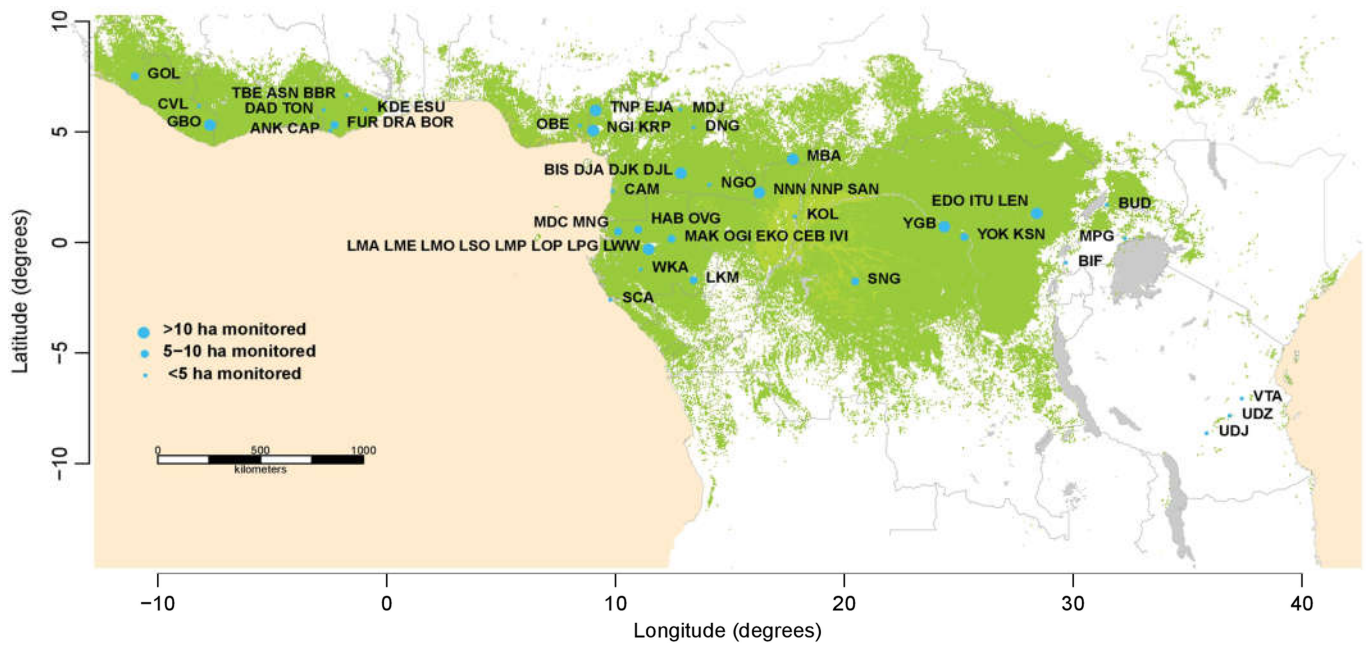
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2035-0>.

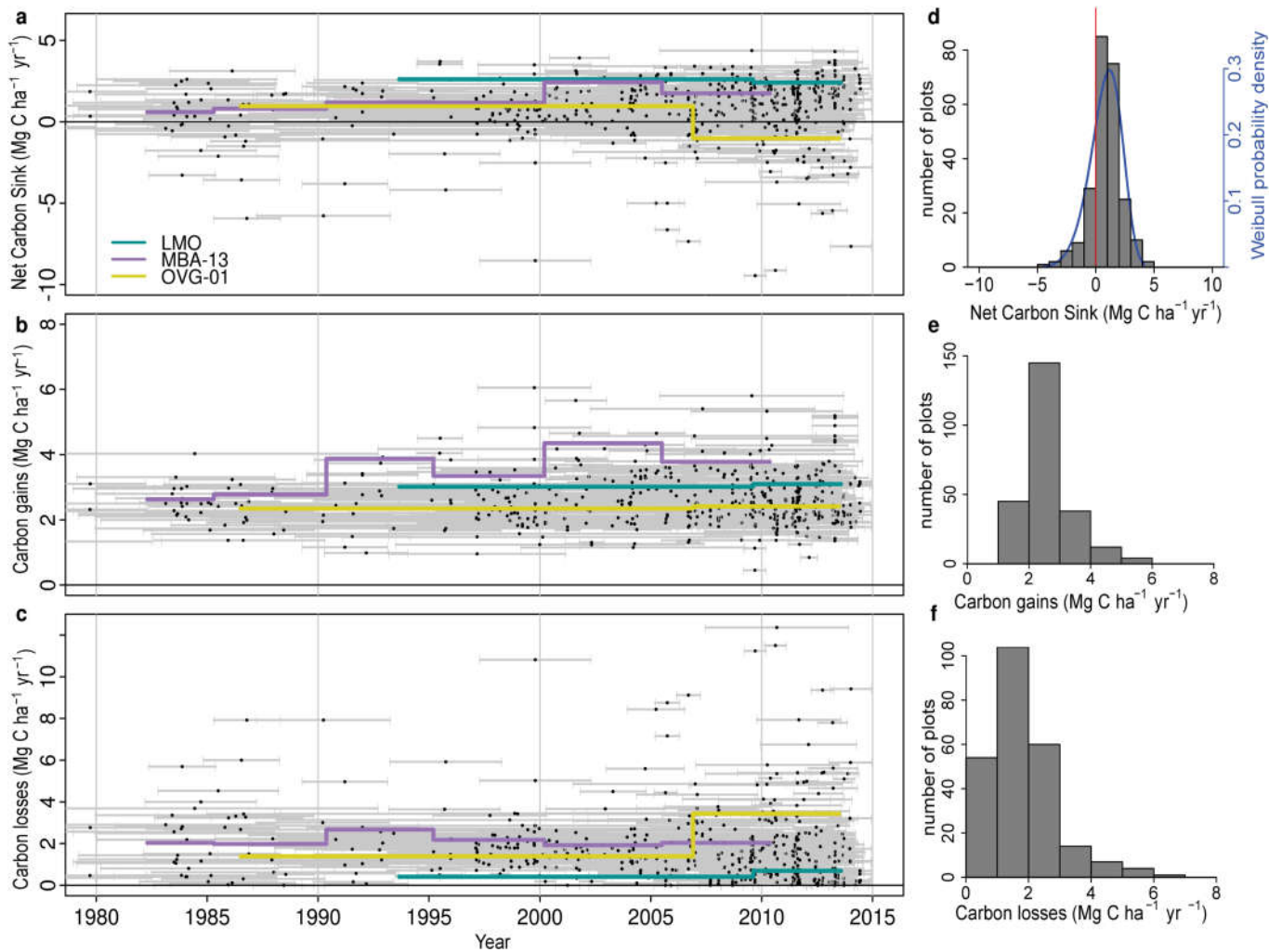
Correspondence and requests for materials should be addressed to W.H.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Map showing the locations of the 244 plots included in this study. Dark green represents all lowland closed-canopy forests, submontane forests and forest-agriculture mosaics; light green shows swamp forests and mangroves, blue circles represent plot clusters, referred to by three-letter codes (see Supplementary Table 1 for the full list of plots). Clusters

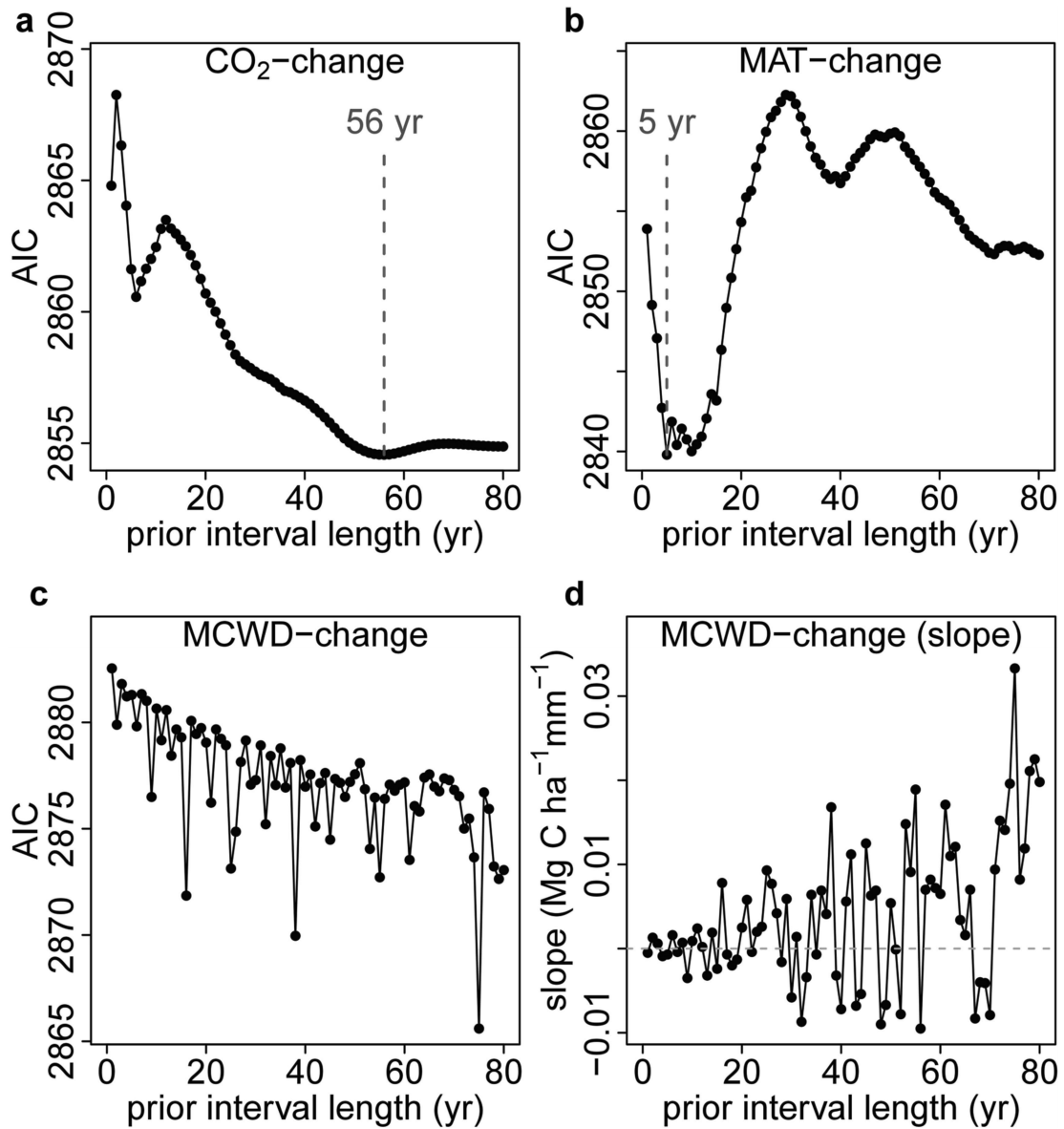
<50 km apart are shown as one point for display only, with the circle size corresponding to sampling effort in terms of hectares monitored. Land cover data are from *The Land Cover Map for Africa in the Year 2000* (GLC2000 database)^{101,102}. This map was created using the R statistical platform, version 3.2.1 (ref.⁶²), which is under the GNU Public License.



Extended Data Fig. 2 | Long-term aboveground carbon dynamics of 244 African structurally intact old-growth tropical forest inventory plots.

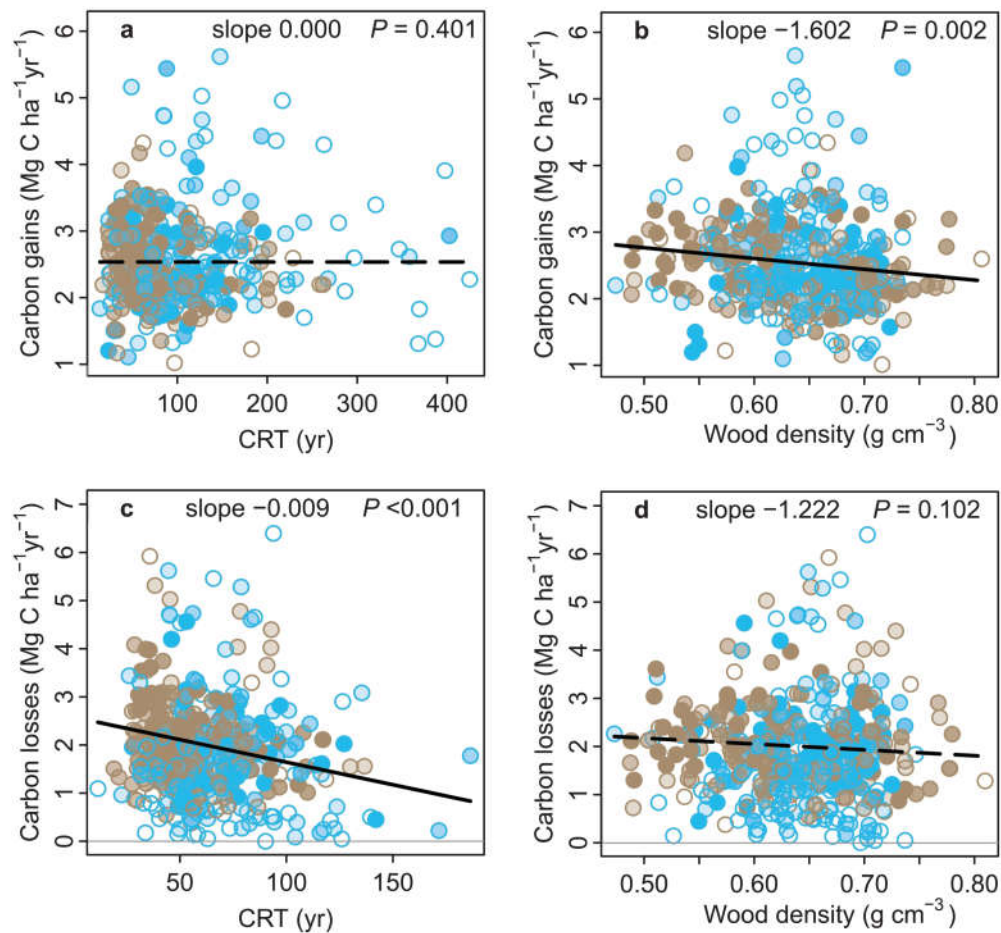
Points in the scatterplots indicate the mid-census interval date, with horizontal bars connecting the start and end date for each census interval for net aboveground biomass carbon change (a), carbon gains (from woody production from tree growth and newly recruited stems) (b), and carbon losses

(from tree mortality) (c). Examples of time series for three individual plots are shown in purple, yellow and green. Associated histograms show the distribution of the plot-level net aboveground biomass carbon (with a three-parameter Weibull probability density distribution fitted in blue, showing that the carbon sink is significantly larger than zero; one-tailed *t*-test: $P < 0.001$) (d), carbon gains (e) and carbon losses (f).



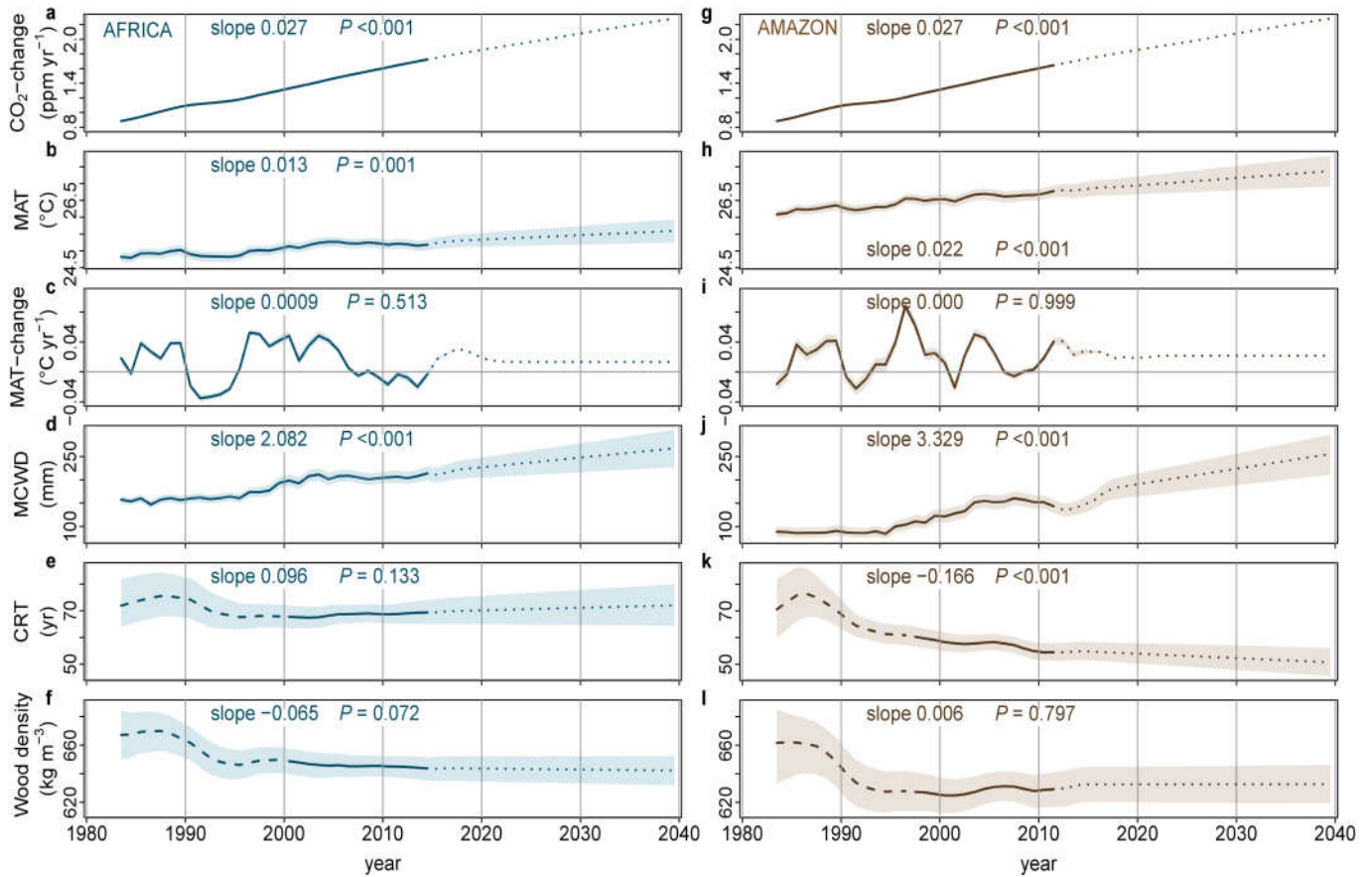
Extended Data Fig. 3 | AIC from correlations between the carbon gain in tropical forest inventory plots and changes in atmospheric CO₂, temperature (MAT) or drought (MCWD), each calculated over ever-longer prior intervals. Panels show the AIC from linear mixed effects models of carbon gains from 565 African and Amazonian plots and corresponding changes in atmospheric CO₂ (CO₂-change) (a), MAT (MAT-change) (b), and drought (MCWD-change) (c). For CO₂ the AIC minimum was observed when predicting the carbon gain from the change in CO₂ calculated over a 56-year-long prior interval length. We use this length of time to calculate our CO₂-change parameter. Such a value is expected because forest stands will respond most strongly to CO₂ when most individuals have grown under the new rapidly changing condition, which should be at its maximum at a time approximately equivalent to the CRT of a forest stand^{30,90} (mean of 62 years in

this pooled African and Amazonian dataset). For MAT the AIC minimum was 5 years, which we use as the prior interval to calculate our MAT-change parameter. This length is consistent with experiments showing temperature acclimation of leaf- and plant-level photosynthetic and respiration processes over approximately half-decadal timescales^{31,91}. For MCWD the AIC minimum is not obvious, while the slope of the correlation, shown in d, shows no overall trend and oscillates between positive or negative values, meaning there is no relationship between carbon gains and the change in MCWD over intervals longer than 1 year; therefore MCWD-change is not included in our models. This result suggests that once a drought ends, its impact on tree growth fades rapidly, as seen in other studies^{14,92}. Furthermore, in the moist tropics wet-season rainfall is expected to recharge soil water, and hence lagged impacts of droughts are not expected.



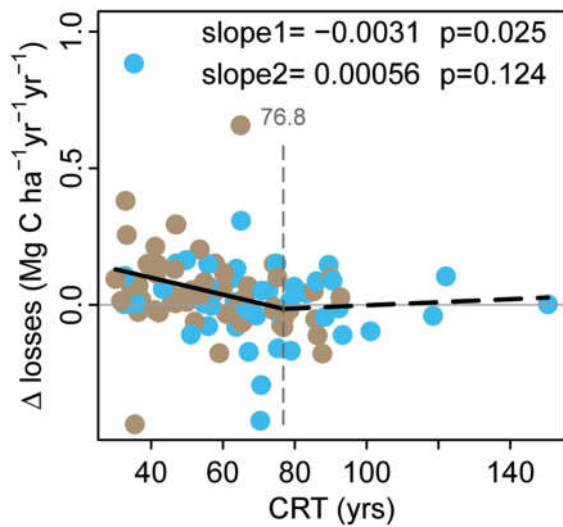
Extended Data Fig. 4 | Potential forest dynamics-related drivers of carbon gains and losses in structurally intact old-growth African and Amazonian tropical forest inventory plots. The aboveground carbon gains, from woody production (a, b), and aboveground carbon losses, from tree mortality (c, d), are plotted against the CRT, and wood density for African (blue) and Amazonian (brown) inventory plots. Linear mixed effects models were performed with census intervals ($n=1,566$) nested within plots ($n=565$) to avoid pseudo-replication, using an empirically derived weighting based on interval length and plot area (see Methods). Significant regression lines from

the linear mixed effects models for the complete dataset are shown as a solid line; non-significant regressions are shown as a dashed line. Each dot represents a time-weighted mean plot-level value; the shading of the dot represents total monitoring length, with empty circles corresponding to plots monitored for ≤ 5 years and solid circles for plots monitored for > 20 years. Carbon loss data are presented untransformed for comparison with carbon gains; linear mixed effects models on transformed data to fit normality assumptions do not change the significance of the results. Note that CRT is calculated differently for the carbon gains and losses models (see Methods).

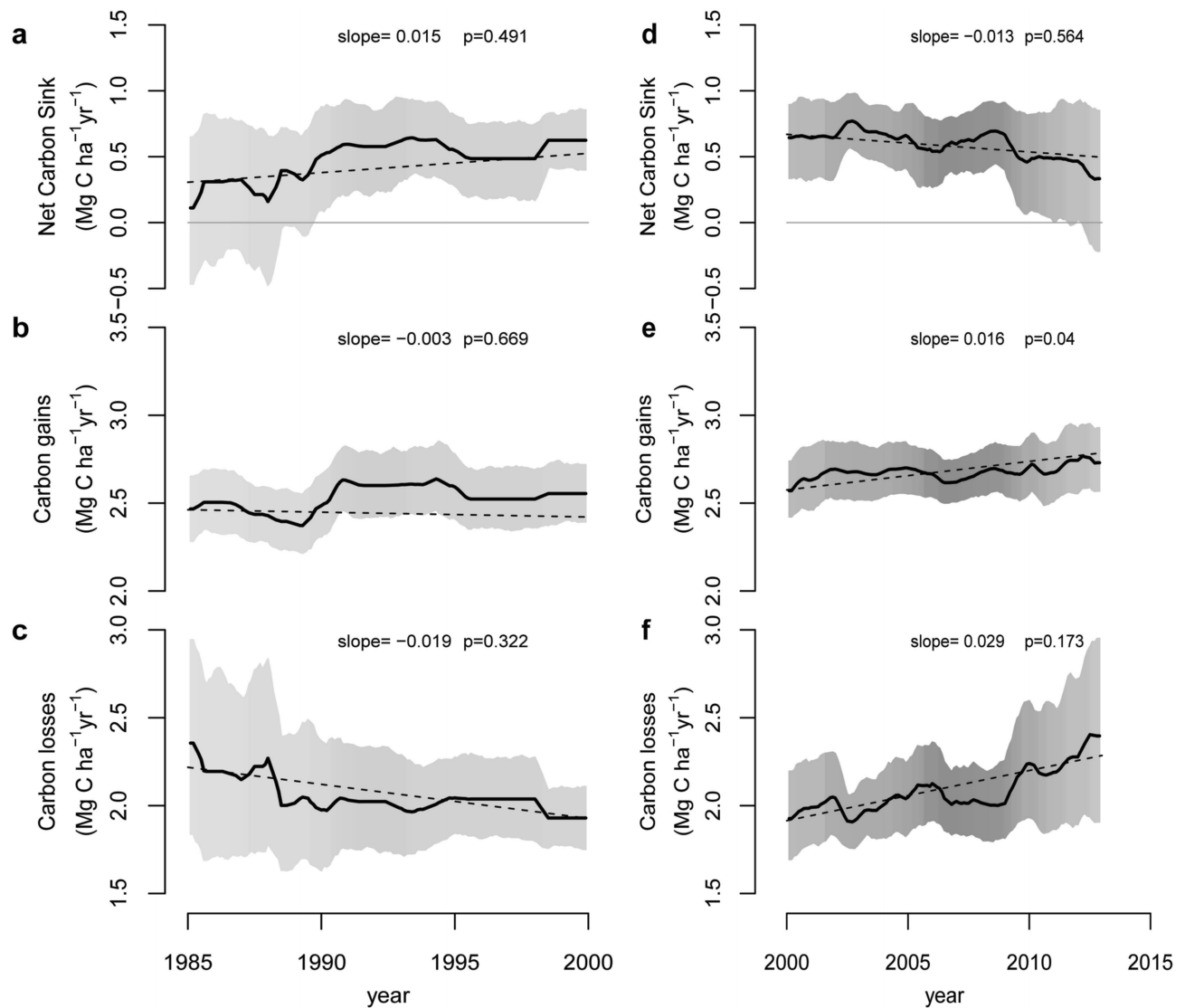


Extended Data Fig. 5 | Trends in predictor variables used to estimate long-term trends in aboveground carbon gains, carbon losses and the resulting net carbon sink in African and Amazonian structurally intact old-growth tropical forest inventory plot networks. Mean annual CO₂-change (a), MAT (b), MAT-change (c), MCWD (d), CRT (e) and wood density (f) for African plot locations in blue, and corresponding variables for Amazon plot locations in brown (g–l). Solid lines represent observational data where >75% of the plots were monitored; long-dashed lines are plot means where <75% of plots were monitored. Dotted lines are future values estimated from linear trends from

the 1 January 1983 to 31 December 2014 (Africa) or 1 January 1983 to mid-2011 (Amazon) data (slope and P value reported in each panel), see Methods for details. Upper and lower confidence intervals (shaded area) for the past are calculated by respectively adding and subtracting 2σ to the mean of each annual value. Upper and lower confidence intervals for the future (Africa: 1 January 2015 to 31 December 2039; Amazonia: mid-2011 to 31 December 2039) were estimated by adding and subtracting 2σ from the slope of the regression model.

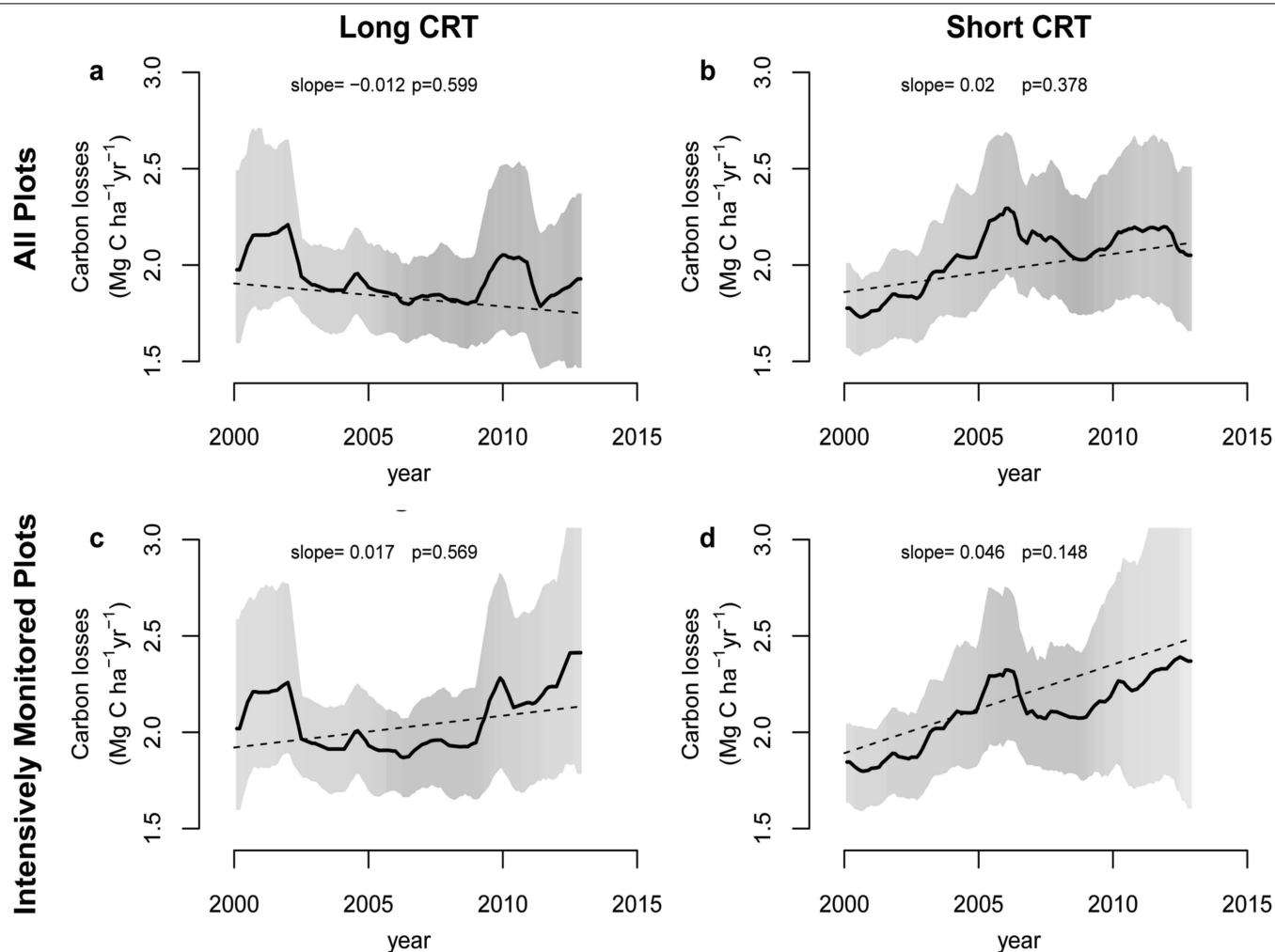


Extended Data Fig. 6 | The change in carbon losses versus CRT of long-term structurally intact old-growth forest inventory plots in Africa and Amazonia. For plots with two census intervals, we calculated the change in carbon losses ('Δlosses') as the carbon losses (in $\text{Mg C ha}^{-1}\text{yr}^{-1}$) of the second interval minus the carbon losses of the first interval, divided by the difference in mid-interval dates. For plots with more than two intervals, we calculated the change in carbon losses for each pair of subsequent intervals, then calculated the plot-level mean over all pairs, weighted by the time length between mid-interval dates. This analysis includes only plots with at least two census intervals that were monitored for a total of ≥ 20 years (that is, roughly one-third of the mean CRT of the pooled African and Amazon dataset; $n = 116$). Breakpoint regression was used to assess the CRT length below which forest carbon losses begin to increase. Plots with $\text{CRT} < 77$ years show a recent long-term increase in carbon losses; longer CRT plots do not. Blue points are African plots, brown points are Amazonian plots.



Extended Data Fig. 7 | Trends in net aboveground live biomass carbon, carbon gains and carbon losses from intensively monitored structurally intact old-growth tropical forest inventory plots in Africa. Trends are calculated for the last 15 years of the twentieth century (a–c) and the first 15 years of the twenty-first century (d–f). Plots were selected from the full dataset if their census intervals cover at least 50% of the respective time windows, that is, they are intensively monitored ($n = 56$ plots for 1 January 1985 to 31 December 1999, and $n = 134$ plots for 1 January 2000 to 31 December 2014, respectively).

Solid lines show mean values, and shading corresponds to the 95% CI, as calculated in Fig. 1. Dashed lines, slopes and P values are from linear mixed effects models, as in Fig. 1. The data shows a difference compared to Fig. 1, notably the sink decline after about 2010 driven by rising carbon losses. This is because in Fig. 1 we include all available plots over the 1 January 1983 to 31 December 2014 window, which includes clusters of plots monitored only in the 2010s, often monitored for a single census interval, that had low carbon loss and high carbon sink values.



Extended Data Fig. 8 | Twenty-first-century trends in aboveground biomass carbon losses from structurally intact old-growth African tropical forest inventory plots with either long or short CRT. a, b, All plots, that is, as in Fig. 1, but split into a long-CRT group (a) and a short-CRT group (b), each containing half of the 244 plots. **c, d,** Plots are restricted to those spanning >50% of the time window, that is, intensively monitored plots, as in Extended Data Fig. 7, but

split into a long-CRT group (c) and a short-CRT group (d), each containing half of the 134 plots. Solid lines indicate mean values, shading the 95% CI, as for Fig. 1. Dashed lines, slopes and P values are from linear mixed effects models, as for Fig. 1. Carbon losses increase at a higher rate in the short-CRT than the long-CRT group of plots, in both datasets, although this increase is not statistically significant.

Extended Data Table 1 | Models to predict carbon gains and losses in structurally intact old-growth African and Amazonian tropical forests

Carbon gains (Mg C ha ⁻¹ yr ⁻¹)				
Predictor variable	Parameter value	Standard error	<i>t</i> value	<i>P</i> value
Intercept	4.694	0.739	6.354	<0.001
CO ₂ (ppm)	0.005	0.001	3.196	0.001
MAT (°C)	-0.143	0.021	-6.844	<0.001
MCWD (mm x1,000)	-1.232	0.210	-5.878	<0.001
Carbon losses (Mg C ha ⁻¹ yr ⁻¹)				
Predictor variable	Parameter value	Standard error	<i>t</i> value	<i>P</i> value
Intercept	0.926	1.854	0.499	0.617
CO ₂ (ppm)	0.004	0.004	0.947	0.344
MAT (°C)	-0.011	0.044	-0.249	0.804
MCWD (mm x1,000)	-0.498	0.505	-0.985	0.325

Models to predict carbon gains and losses in structurally intact old-growth African and Amazonian tropical forests, including only environmental variables that show long-term trends that affect theory-driven models of photosynthesis and respiration. Carbon loss values were normalized via power-law transformation, $\lambda = 0.361$.

Extended Data Table 2 | Forest area estimates used to calculate total continental forest sink

Period	intact forest area (Mha)			
	Africa	Amazon	Southeast Asia	Pan-tropics
1980	671.5	958.3	233.6	1863.4
1985	634.3	921.1	207.4	1762.8
1990	600.2	885.2	190.6	1676.0
1995	565.9	851.1	163.5	1580.5
2000	531.8	817.2	136.9	1485.9
2005	504.8	784.5	129.2	1418.5
2010	477.8	756.3	118.4	1352.5
2015	450.5	726.7	101.5	1278.7
2020	425.5	698.5	90.1	1214.2
2025	402.0	671.5	80.0	1153.4
2030	379.7	645.4	71.0	1096.1
2035	358.6	620.4	63.0	1042.1
2040	338.8	596.4	56.0	991.1

Intact forest area for 1990, 2000, 2005 and 2010 is published in ref. ¹ (that is, the total forest area minus forest regrowth). To estimate intact forest area for the other years in this table, we fitted exponential models for each continent using the published data¹.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection.

Data analysis

All calculations were performed using the R statistical platform, version 3.2.1 (R Development Core Team, 2015) using the BiomasaFP R package v0.2.1 (Lopez-Gonzalez, Sullivan, & Baker, 2017). Source data and R-code to generate figures and tables are available from: <https://figshare.com/s/60f48673202283421f43>.

References:

Lopez-Gonzalez, G., Sullivan, M., & Baker, T. 2017. BiomasaFP package. Tools for analysing data downloaded from ForestPlots.net. R package version 0.2.1. Available at <http://www.forestplots.net/en/resources/analysis>.
R Development Core Team. 2015. R: A Language and Environment for Statistical Computing. Available at <http://www.R-project.org/>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data and R-code to generate figures and tables are available from: <https://figshare.com/s/60f48673202283421f43>. This data and code package allows

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We reconstruct the evolution of the per unit area African tropical forest carbon sink (in Mg C ha⁻¹ yr⁻¹) over three decades to 2015 (Figure 1). To do so, we collected, compiled and analysed data from 244 repeatedly measured permanent forest inventory plots in 11 African countries. Selected plots are situated in structurally intact old-growth forests and are part of the African Tropical Rainforest Observation Network (AfriTRON; www.afritron.org; 217 plots) and other sources (27 plots). Plot monitoring periods span 2 to 40 years, between 1968 to 2015 (Extended Data Figure 1). In each plot (mean size, 1.1 ha), all trees ≥100 mm in stem diameter were identified, mapped and measured on at least two occasions using standardized methods (135,625 trees monitored) and live biomass carbon stocks were estimated for each census date, with carbon gains and losses calculated for each interval (Extended Data Figure 2). We compared trends in the per unit area African tropical forest carbon sink with published long-term trends in the Amazonian carbon sink (Brienen, et al. 2015). We pooled the new African and existing Amazonian plot inventory data together to investigate the putative environmental drivers of changes in the tropical forest carbon sink, and project its likely future evolution.

Aboveground Carbon (AGC, in Mg C ha⁻¹) in living biomass for each plot at each census date was estimated as the sum of the AGC of each living stem, then divided by plot area (in hectares).

Carbon Gain is the sum of the aboveground live biomass carbon additions from the growth of surviving stems and the addition of newly recruited stems, using standard methods (Brienen, et al. 2015). For each stem that survived a census interval, carbon additions from its growth (Mg C ha⁻¹ yr⁻¹) were calculated as the difference between its AGC at the end census of the interval and its AGC at the beginning census of the interval. For each stem that recruited during the census interval (i.e. reaching DBH≥100 mm), carbon additions were calculated in the same way, assuming DBH=0 mm at the start of the interval (Talbot, et al. 2014). The carbon additions in an interval, from surviving and newly recruited stems, were summed, then divided by the census interval length (in years), and scaled by plot area (in hectares) (Talbot, et al. 2014). As carbon gains are affected by a census interval bias, with the underestimate increasing with census length, we corrected this bias by accounting for (i) the carbon additions from trees that recruited and then died within the same interval (unobserved recruitment), and (ii) the carbon additions from trees that grew before they died within an interval (unobserved growth) (Talbot, et al. 2014). These typically add <3% to plot-level carbon gains.

Carbon Loss (in Mg C ha⁻¹ yr⁻¹) is estimated, using standard methods (Brienen, et al. 2015), as the sum of aboveground biomass carbon from all stems that died during a census interval, divided by the census length (in years) and scaled by plot area (in hectares). Carbon loss is also affected by the same census interval bias, hence we corrected this bias by accounting for (i) the additional carbon losses from the trees that were recruited and then died within the same interval, and (ii) the additional carbon losses resulting from the growth of the trees that died in the interval (Kohyama, et al. 2018; Talbot, et al. 2014). Calculation details of both components are explained in Supplementary Methods.

Net Carbon Sink (in Mg C ha⁻¹ yr⁻¹) is estimated as carbon gains minus carbon losses.

The estimated mean carbon gains, carbon losses and the net carbon sink of the African plots from 1983-2014, the solid lines in Figure 1, were calculated following (Brienen, et al. 2015) to allow direct comparison with published Amazonian results. First, each census interval value was interpolated for each 0.1-yr period within the census interval. Then, for each 0.1-yr period between 1983 and 2014, we calculate a weighted mean of all plots monitored at that time, using the square root of plot area as a weighting factor. Finally, confidence intervals for each 0.1-yr period are bootstrapped.

References:

Brienen, R. J. W., et al.
2015 Long-term decline of the Amazon carbon sink. *Nature* 519(7543):344-348.
Kohyama, Takashi S., et al.
2018 Definition and estimation of vital rates from repeated censuses: Choices, comparisons and bias corrections focusing on trees. *Methods in Ecology and Evolution* 9(4):809-821.
Talbot, Joey, et al.
2014 Methods to estimate aboveground wood productivity from long-term forest inventory plots. *Forest Ecology and Management* 320:30-38.

Research sample

We use data from 244 plots in 11 African countries to present the first assessment of the temporal evolution of the tropical forest carbon sink in Africa. It represents 10 years of new field campaigns in Africa, extending sampling into extremely remote and previously unsampled regions. This is the first new manuscript using long-term inventory plots to estimate the intact forest carbon sink in Africa since (Lewis, et al. 2009) was published in *Nature*.
Plot selection: 244 permanent inventory plots were selected from 11 countries. These plots are situated in closed canopy (i.e. not

woody savanna) old-growth mixed-age forests and were selected using commonly used criteria (Brienen, et al. 2015; Lewis, et al. 2009; Lewis, et al. 2013): free of fire and industrial logging; all trees with diameter at reference height ≥ 100 mm measured at least twice; ≥ 0.2 ha area; < 1500 m.a.s.l. altitude; MAT ≥ 20.0 °C (Hijmans, et al. 2005); annual precipitation ≥ 1000 mm; located ≥ 50 m from anthropogenic forest edges.

References:

Brienen, R. J. W., et al.
2015 Long-term decline of the Amazon carbon sink. *Nature* 519(7543):344-348.
Hijmans, Robert J., et al.
2005 Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25(15):1965-1978.
Lewis, S. L., et al.
2009 Increasing carbon storage in intact African tropical forests. *Nature* 457(7232):1003-1006.
Lewis, Simon L., et al.
2013 Above-ground biomass and structure of 260 African tropical forests. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1625):20120295-20120295.

Sampling strategy

No sample size calculation was performed. We selected all available plots meeting the criteria described above. All African tropical forest regions (West Africa, Lower Guinea, Congo Basin, East Africa) are adequately represented. This is the largest dataset of repeatedly measured plots ever used to calculate long-term trends in African forest carbon dynamics.

Data collection

Plot inventory data was collected by teams led by at least one of the 104 researchers co-authoring this paper. All permanent inventory plots are part of one or several networks. Of the 244 plots included in the study, 217 contribute to the African Tropical Rainforest Observatory Network (AfriTRON; www.afritron.org), with data curated at www.ForestPlots.net. These include plots from Sierra Leone, Liberia, Ghana, Nigeria, Cameroon, Gabon, Republic of Congo, Democratic Republic of Congo (DRC), Uganda and Tanzania (Lopez-Gonzalez, et al. 2011; Lopez-Gonzalez, et al. 2009) (Extended Data Figure 1). Fifteen plots are part of the TEAM network, from Cameroon, Republic of Congo, Tanzania, and Uganda (Hockemba 2010; Kenfack 2011; Rovero, et al. 2009; Sheil and Bitariho 2009). Nine plots contribute to the ForestGEO network, from Cameroon and DRC (Anderson-Teixeira, et al. 2015) (9 plots from DRC, codes SNG, contribute to both AfriTRON and ForestGEO networks, included above in the AfriTRON total). Finally, three plots from Central African Republic are part of the CIRAD network (Claeys, et al. 2019; Gourlet-Fleury, et al. 2013).

Tree-level aboveground biomass carbon is estimated using an allometric equation (Chave, et al. 2014) with parameters for tree diameter, tree height and wood mass density. The estimated aboveground biomass of a plot is the sum of the estimated biomass of all live trees at that census date.

Tree Diameter: In all plots, all woody stems with ≥ 100 mm diameter at 1.3 m from the base of the stem ('diameter at breast height', DBH), or 0.5 m above deformities or buttresses, were measured, mapped and identified using standard forest inventory methods (Phillips, et al. 2016). The height of the point of measurement (POM) was marked on the trees and recorded, so that the same POM is used at the subsequent forest census. For stems developing deformities or buttresses over time that could potentially disturb the initial POM, the POM was raised approximately 500 mm above the deformity. Estimates of the diameter growth of trees with changed POM used the ratio of new and old POMs, to create a single trajectory of growth from the series of diameters at two POM heights (Brienen, et al. 2015; Lewis, et al. 2009; Talbot, et al. 2014). We used standardized protocols to assess typographical errors and potentially erroneous diameter values (e.g. trees shrinking by >5 mm), missing values, failures to find the original POM, and other issues. Where necessary we estimated the likely value via interpolation or extrapolation from other measurements of that tree, or when this was not possible we used the median growth rate of trees in the same plot, census and size-class, defined as DBH = 100-199 mm, or 200-399 mm, or >400 mm (Talbot, et al. 2014). We interpolate measurements for 1.3% of diameters, extrapolate 0.9%, and use median growth rates for 1.5%.

Tree height: Height of individuals from ground to the top leaf, hereafter H_t , was measured in 204 plots, using a laser hypsometer (Nikon forestry Pro) from directly below the crown (most plots), a laser or ultrasonic distance device with an electronic tilt sensor, a manual clinometer, or by direct measurement, i.e. tree climbing. Only trees where the top was visible were selected (Sullivan, et al. 2018). In most plots, tree selection was similar: the 10 largest trees were measured, together with 10 randomly selected trees per diameter from five classes: 100-199 mm, 200-299 mm, 300-399 mm, 400-499 mm, and 500+ mm trees, following standard protocols (Sullivan, et al. 2018). We use these data and the local heights function in R package BiomasaFP (Lopez-Gonzalez, et al. 2017) to fit 3-parameter Weibull relationships (see Supplementary Methods for a full explanation of this procedure):

$$H_t = a \times (1 - e^{(-b \times (DBH/10)^c)}) \quad (\text{equation 1}).$$

We chose the Weibull model as it is known to be robust when a large number of measurements are available (Feldpausch, et al. 2012; Sullivan, et al. 2018). We parameterize this H_t -DBH relationship for four different combinations of edaphic forest type and biogeographical region (parameters in parentheses): (i) terra firme forest in West Africa ($a=56.0$; $b=0.0401$; $c=0.744$); (ii) terra firme forest in Lower Guinea and Western Congo Basin ($a=47.6$; $b=0.0536$; $c=0.755$); (iii) terra firme forest in Eastern Congo Basin and East Africa ($a=50.8$; $b=0.0499$; $c=0.706$); and finally (iv) seasonally flooded forest from Lower Guinea and Western Congo Basin ($a=38.2$; $b=0.0605$; $c=0.760$). The parameters were used to estimate H_t from DBH for all tree DBH measurements for input into the allometric equation.

Wood Density: Dry wood density (ρ) measurements were compiled for 730 African species from published sources and stored in www.ForestPlots.net; most were sourced from the Global Wood Density Database on the Dryad digital repository (www.datadryad.org) (Chave, et al. 2009; Zanne, et al. 2009). Each individual in the tree inventory database was matched to a species-specific mean wood density value. Species in both the tree inventory and wood density databases were standardized for orthography and synonymy using the African Flowering Plants Database (www.ville-ge.ch/cjb/bd/africa/) to maximize matches (Lewis, et al. 2009). For incompletely identified individuals or for individuals belonging to species not in the ρ database, we used the mean ρ value for the next higher known taxonomic category (genus or family, as appropriate). For unidentified individuals, we used the mean wood density value of all individual trees in the plot (Lewis, et al. 2009; Lopez-Gonzalez, et al. 2011).

Allometric equation: For each tree we use a published allometric equation (Chave, et al. 2014) to estimate aboveground biomass. We

then convert this to carbon, assuming that aboveground carbon (AGC) is 45.6% of aboveground biomass (Martin, et al. 2018). Thus: $AGC = 0.456 \times (((0.0673 \times (\rho \times (DBH/10)^2 \times H_t)^{0.976}) / (1000)))$ (equation 2), with DBH in mm, dry wood density, ρ , in g cm⁻³, and total tree height, H_t , in m (Chave, et al. 2014).

References:

- Anderson-Teixeira, Kristina J., et al.
2015 CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Global Change Biology* 21(2):528-549.
- Brienen, R. J. W., et al.
2015 Long-term decline of the Amazon carbon sink. *Nature* 519(7543):344-348.
- Chave, Jerome, et al.
2009 Towards a worldwide wood economics spectrum. *Ecology Letters* 12(4):351-366.
- Chave, Jérôme, et al.
2014 Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20(10):3177-3190.
- Claeys, Florian, et al.
2019 Climate change would lead to a sharp acceleration of Central African forests dynamics by the end of the century. *Environmental Research Letters* 14(4):044002.
- Feldpausch, T. R., et al.
2012 Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* 9(8):3381-3403.
- Gourlet-Fleury, Sylvie, et al.
2013 Tropical forest recovery from logging: a 24 year silvicultural experiment from Central Africa. *Philosophical Transactions of the Royal Society B-Biological Sciences* 368(1625):20120302.
- Hockemba, M. B. N.
2010 Nouabalé Ndoki TEAM Site. Data Set Identifier: TEAM-DataPackage-20151201235855_1254.
- Kenfack, D.
2011 Korup National Park TEAM Site. Data Set Identifier: TEAM-DataPackage-20151201235855_1254.
- Lewis, S. L., et al.
2009 Increasing carbon storage in intact African tropical forests. *Nature* 457(7232):1003-1006.
- Lopez-Gonzalez, G., et al.
2011 ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. *Journal of Vegetation Science* 22:610–613.
- Lopez-Gonzalez, G., et al.
2009 ForestPlots.net Database. www.forestplots.net. Date of extraction [10/11/2017].
- Lopez-Gonzalez, Gabriela, Martin Sullivan, and Tim Baker
2017 BiomasaFP package. Tools for analysing data downloaded from ForestPlots.net. R package version 0.2.1. Available at <http://www.forestplots.net/en/resources/analysis>.
- Martin, Adam R., Mahendra Doraisami, and Sean C. Thomas
2018 Global patterns in wood carbon concentration across the world's trees and forests. *Nature Geoscience* 11(12):915-920.
- Phillips, O., et al.
2016 RAINFOR field manual for plot establishment and remeasurement. Available at http://www.rainfor.org/upload/ManualsEnglish/RAINFOR_field_manual_version_2016.pdf.
- Rovero, F., A. Marshall, and E. Martin
2009 Udzungwa TEAM Site. Data Set Identifier: TEAM-DataPackage-20151130235007_5069.
- Sheil, D., and R. Bitariho
2009 Bwindi Impenetrable Forest TEAM Site. Data Set Identifier: TEAM-DataPackage-20151201235855_1254.
- Sullivan, M. J. P., et al.
2018 Field methods for sampling tree height for tropical forest biomass estimation. *Methods in Ecology and Evolution* 9(5):1179-1189.
- Talbot, Joey, et al.
2014 Methods to estimate aboveground wood productivity from long-term forest inventory plots. *Forest Ecology and Management* 320:30-38.
- Zanne, A.E., et al.
2009. Data from: Towards a worldwide wood economics spectrum: Dryad Digital Repository.

Timing and spatial scale	<p>The large majority of plots are sited in terra firme forests and have mixed species composition, although four are in seasonally flooded forest and 14 plots are in Gilbertiodendron dewevrei monodominant forest, a locally common forest type in Africa (Supplementary Table 1). The 244 plots have a mean size of 1.1 ha (median, 1 ha), with a total plot area of 277.9 ha. The dataset comprises 391,968 diameter measurements on 135,625 stems, of which 89.9% were identified to species, 97.5% to genus and 97.8% to family.</p> <p>Plots were measured at least twice and maximum 10 times, between 1968 and 2015. Plot monitoring periods span 2 to 40 years; mean total monitoring period is 11.8 years, mean census length 5.7 years, with a total of 3,214 ha years of monitoring. The 321 Amazon plots are published and were selected using the same criteria (ref.6), (Brienen, et al. 2015) except in the African selection criteria we specified a minimum anthropogenic edge distance and added a minimum temperature threshold.</p> <p>Brienen, R. J. W., et al. 2015 Long-term decline of the Amazon carbon sink. <i>Nature</i> 519(7543):344-348.</p>
Data exclusions	Plots were selected using the criteria described above (section Research sample). Plots that did not meet one or several of these criteria were not used for analysis.
Reproducibility	Our analysis does not include experimental findings.
Randomization	Trends in carbon gains, losses and the net carbon sink over time were assessed using linear mixed effects models (lmer function in R,

Randomization

lme4 package (Bates, et al. 2013)), providing the linear slopes reported in Figure 1. These models regress the mid-point of each census interval against the value of the response variable for that census interval. Plot identity was included as a random effect, i.e. assuming that the intercept can vary randomly among plots. Observations were weighted by plot size and census interval length. Weightings were derived empirically, by assuming a priori that there is no significant relation between the net carbon sink and census interval length or plot size (Lewis, et al. 2009).

References:

Bates, D., et al.
2013 lme4: Linear mixed-effects models using Eigen and S4. R package version, 1.0-4. Available at <http://www.inside-r.org/packages/lme4/versions/1-0-4>.
Lewis, S. L., et al.
2009 Increasing carbon storage in intact African tropical forests. *Nature* 457(7232):1003-1006.

Blinding

Blinding was not relevant to our study.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions

All plots are located in African tropical forests receiving at least 1000 mm rainfall annually and with a mean annual temperature of at least 20 °C.

Location

Plots are located at low elevations (<1500 m.a.s.l. altitude). A map showing locations of all plots is presented in Extended Data Figure 1.

Access and import/export

This paper is a product of the African Tropical Rainforest Observatory Network (AfriTRON), the TEAM network, the ForestGEO network, and the CIRAD network. These permanent inventory plot networks only exist thanks to the support of governments, local administrations and villages across Africa who have given us permission for, and helped us complete, our fieldwork. A full list of partner institutions (excluding those in the co-author affiliations) can be found in (on-line only) acknowledgements. Furthermore, plot inventory data are the product of many field-teams which mainly consisted of local assistants. A full list of people involved in data collection can be found in (on-line only) acknowledgements, along with a full list of villages and communities that hosted the field-teams and provided logistical and infrastructural support.

This paper includes 264 plot-censuses (out of 746) that are published for the first time here, including censuses from plots located in extremely remote areas such as the Salonga National Park in the heart of the Congo Basin. Each plot-census represents several months of preparation, transport, data collection, digitalisation and data quality assessment.

Disturbance

No significant disturbance was caused by our measurements. Trees were tagged using a single aluminum nail (no iron), avoiding damage to trees due to corrosion.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Intraplate volcanism originating from upwelling hydrous mantle transition zone


<https://doi.org/10.1038/s41586-020-2045-y>

Jianfeng Yang¹ & Manuele Faccenda¹✉

Received: 19 June 2019

Accepted: 2 December 2019

Published online: 26 February 2020

 Check for updates

Most magmatism occurring on Earth is conventionally attributed to passive mantle upwelling at mid-ocean ridges, to slab devolatilization at subduction zones, or to mantle plumes. However, the widespread Cenozoic intraplate volcanism in northeast China^{1–3} and the young petit-spot volcanoes^{4–7} offshore of the Japan Trench cannot readily be associated with any of these mechanisms. In addition, the mantle beneath these types of volcanism is characterized by zones of anomalously low seismic velocity above and below the transition zone^{8–12} (a mantle level located at depths between 410 and 660 kilometres). A comprehensive interpretation of these phenomena is lacking. Here we show that most (or possibly all) of the intraplate and petit-spot volcanism and low-velocity zones around the Japanese subduction zone can be explained by the Cenozoic interaction of the subducting Pacific slab with a hydrous mantle transition zone. Numerical modelling indicates that 0.2 to 0.3 weight per cent of water dissolved in mantle minerals that are driven out from the transition zone in response to subduction and retreat of a tectonic plate is sufficient to reproduce the observations. This suggests that a critical amount of water may have accumulated in the transition zone around this subduction zone, as well as in others of the Tethyan tectonic belt¹³ that are characterized by intraplate or petit-spot volcanism and low-velocity zones in the underlying mantle.

The Cenozoic intraplate volcanism in northeast China is located more than 1,000 km westward of the Japan Trench¹, while the young alkaline basalts (0–6 Ma) known as petit-spots outcrop up to 600 km eastward of the trench⁴ (Fig. 1). The formation mechanism of these types of onshore and offshore volcanism is still debated, as there is no geological and geophysical correlation with mantle plumes or arc volcanism^{6,14}.

Seismic tomography models indicate that in this region the Pacific Plate is currently stagnant in the mantle transition zone (MTZ), extending continuously up to nearly 1,000 km to the inland of northeast China^{3,8,10}. Thus, it has been proposed that the Cenozoic intraplate magmatism is related to the dehydration of the Pacific slab in the MTZ^{2,15}.

The primary petit-spot magma has been determined to be volatile-rich with extremely enriched mantle (EM1-like) isotopic compositions^{6,7}. The lack of hotspot tracks in this region excludes a contribution from a mantle plume. It has been postulated that the petit-spot magma forms in the asthenosphere and migrates upward through the oceanic lithosphere by reactive porous flow in response to plate flexure^{4,6,7}.

Based on electrical conductivity surveys, the MTZ probably holds about 0.1 wt% water¹⁶. The MTZ below northeast China and Japan is particularly wet, with at least 0.5–1 wt% water¹⁷. The MTZ is primarily composed of wadsleyite and ringwoodite minerals that can accommodate 1–3 wt% water, which is 1 to 2 orders of magnitude higher than the water (hydrogen) solubility in upper- and lower-mantle minerals. Given the large contrast in water solubility between the MTZ and upper/lower mantle, it is reasonable to expect deep dehydration melting when subducting slabs excite vertical flow in the nearby wet MTZ¹⁸. Indeed, seismic low-velocity zones (LVZs) above 410 km and below 660 km have

been observed not only in Japan^{8–12,14,15}, but also around subduction zones in Europe¹⁹ and the western United States^{20,21}.

To test this hypothesis, we construct two-dimensional numerical experiments in which a self-sustained oceanic plate subduction is characterized by trench retreat and slab stagnation into a homogeneously or heterogeneously wet MTZ (see Methods). The subducting plate and entrained dry upper mantle push the adjacent wet MTZ downward to the lower mantle such that a partially molten layer forms between 700 km and 800 km depth (Fig. 2a, region labelled M2) (Supplementary Video 1). On the other hand, MTZ material uplifted to the upper mantle starts to partially melt above 410 km (Fig. 2a, M3). Slab stagnation and retreat is accompanied by sub-slab MTZ upwelling and new melting (M1). These partially molten regions above and below the MTZ cause large seismic LVZs (Fig. 2c). When melt percolation is active (see Methods), extraction to the surface occurs, forming intraplate and petit-spot volcanisms ahead of and behind the trench, respectively.

Figure 3 shows the spatial and temporal trend of modelled volcanics for the reference model in Fig. 2. The first intraplate volcanism occurs about 500 km away from the trench, then spreads in two opposite directions. The mantle water content decreases after melt extraction, which precludes further deep (≥ 200 km) melting of the residual peridotite (Fig. 2b). As the slab rolls back, more distal wet MTZ is sucked into the upper mantle wedge, such that partial melting and volcanoes will form further away from the trench. The new generated volcanism is not homogeneously distributed as it is strongly influenced by mantle flow and trench movement. Furthermore, a heterogeneous distribution of water in the MTZ would prevent the formation of any

¹Dipartimento di Geoscienze, Università di Padova, Padova, Italy. ✉e-mail: manuele.faccenda@unipd.it

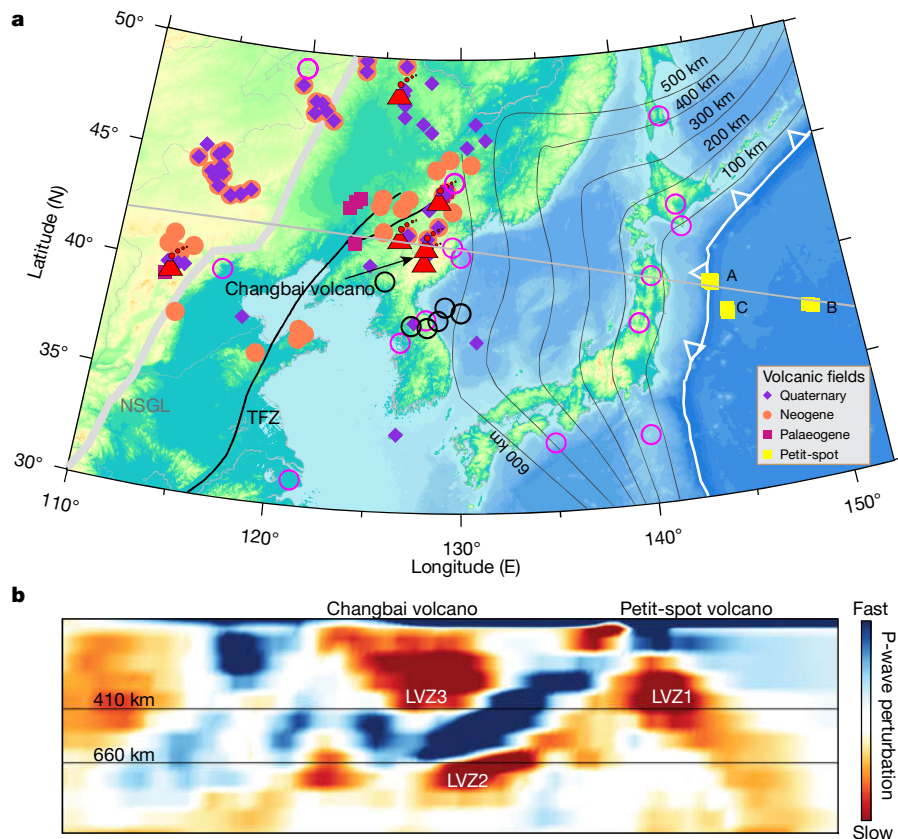


Fig. 1 | Geological/geophysical maps and Cenozoic volcanic fields in northeast China and offshore Japan. a, The red triangles denote volcanoes; the black and magenta open circles show seismic low velocity at 410 km and below 660 km, respectively, as determined by receiver functions^{9,11,12}. The yellow squares indicate three young alkaline basalt sites (A, B, C), known as petit-spot, offshore the Japan Trench. The black contour lines indicate the

Pacific Plate depths in the mantle. The present-day Pacific Plate front lies between the Tanlu fault zone (TFZ) and the north-south gravity lineament (NSGL). Map created with open software GMT 5.4.3. **b**, Cross-section (thick grey line in **a**) of seismic P-wave velocity perturbation with three distinct low-velocity zones⁸.

temporal-spatial magmatic sequence as wetter portions would melt earlier than drier regions at the same pressure and temperature (P - T) conditions. It is noteworthy that intraplate volcanism also occurs a few hundred kilometres in front of the slab tip. After about 12 Myr of modelled subduction, petit-spot volcanoes appear behind the trench. They are located up to about 300 km seaward of the trench and exhibit a similar magmatic activity trend to the intraplate volcanism.

We further test the influence of initial water content in the transition zone and other parameters on the genesis of asthenospheric melting (Extended Data Fig. 5). Melting commences 40 km above the transition zone, and no petit-spot volcanoes are formed for 0.2 wt% initial water. The thickness of the partially molten layer could range from a few tens of kilometres to more than 100 km, depending on the melt extraction efficiency and water content. A petit-spot volcano might be located more than 600 km from the trench if the melt extraction process is efficient (Extended Data Fig. 5b). Given the assumed homogeneous distribution of water in the MTZ, these models provide upper-bound estimates on the volumes of volcanics and melt. However, the results also hold for a more realistic heterogeneous distribution of the water in this mantle level (Extended Data Fig. 5e).

When comparing the model results with seismic and geological observations, we note that around the Pacific slab, three remarkable seismic LVZs outside the transition zone are clearly imaged (Fig. 1b). These are well correlated with the locations of intraplate and petit-spot volcanoes, and the modelled partially molten zones (Figs. 1a and 2). Although seismic low-velocity anomalies are generally attributed to thermal effects¹⁴, or to the presence of water²², melt^{11,20} and/or major element compositional heterogeneities^{16,23}, it has recently been argued

that some of these LVZs could be artefacts induced by seismic anisotropy²⁴. Nevertheless, the authenticity of the sub-slabs LVZ1 and LVZ2 appearing in tomographic models has been confirmed by other independent studies using, respectively, an accurate scrutiny of the seismic ray paths²⁵ sampling the LVZ1, and receiver functions in the case of LVZ2¹¹. The LVZ3 sits below the active Changbai volcano and appears to extend down to 410 km as revealed by multiple high-resolution tomography models^{3,8,15,26}. A thermal anomaly from a non-hotspot upwelling, if it hypothetically exists, is difficult to reconcile with the large velocity drop of LVZ1. The hot material will rapidly cool when flowing upward, because of adiabatic decompression and the latent heat of the wadsleyite-to-olivine reaction. Laboratory experiments show that seismic wave speeds are insensitive to moderate (<1 wt%) water contents for olivine²⁷ and wadsleyite²²; thus, the LVZs are very likely to be caused by partial melting and/or compositional heterogeneities. The presence of basalts at the bottom of the upper mantle can be excluded as it would generate a positive seismic anomaly²⁸. On the other hand, basalts accumulating at the base of the MTZ could be effectively dragged by the slab into the uppermost lower mantle and generate the LVZ2. However, receiver functions indicate that the lower-mantle LVZs are within the 750–780 km depth range¹¹, which is likely to be below the post-garnet phase transition zone where basalts are seismically faster than mantle rocks.

The presence of melt in the deep mantle, which is mostly catalysed by the involvement of volatiles, decreases seismic velocities and provides a magmatic source for intraplate/petit-spot volcanism. Our numerical models thus suggest that a hydrous transition zone with at least 0.2–0.3 wt% water beneath northeast China and offshore Japan can

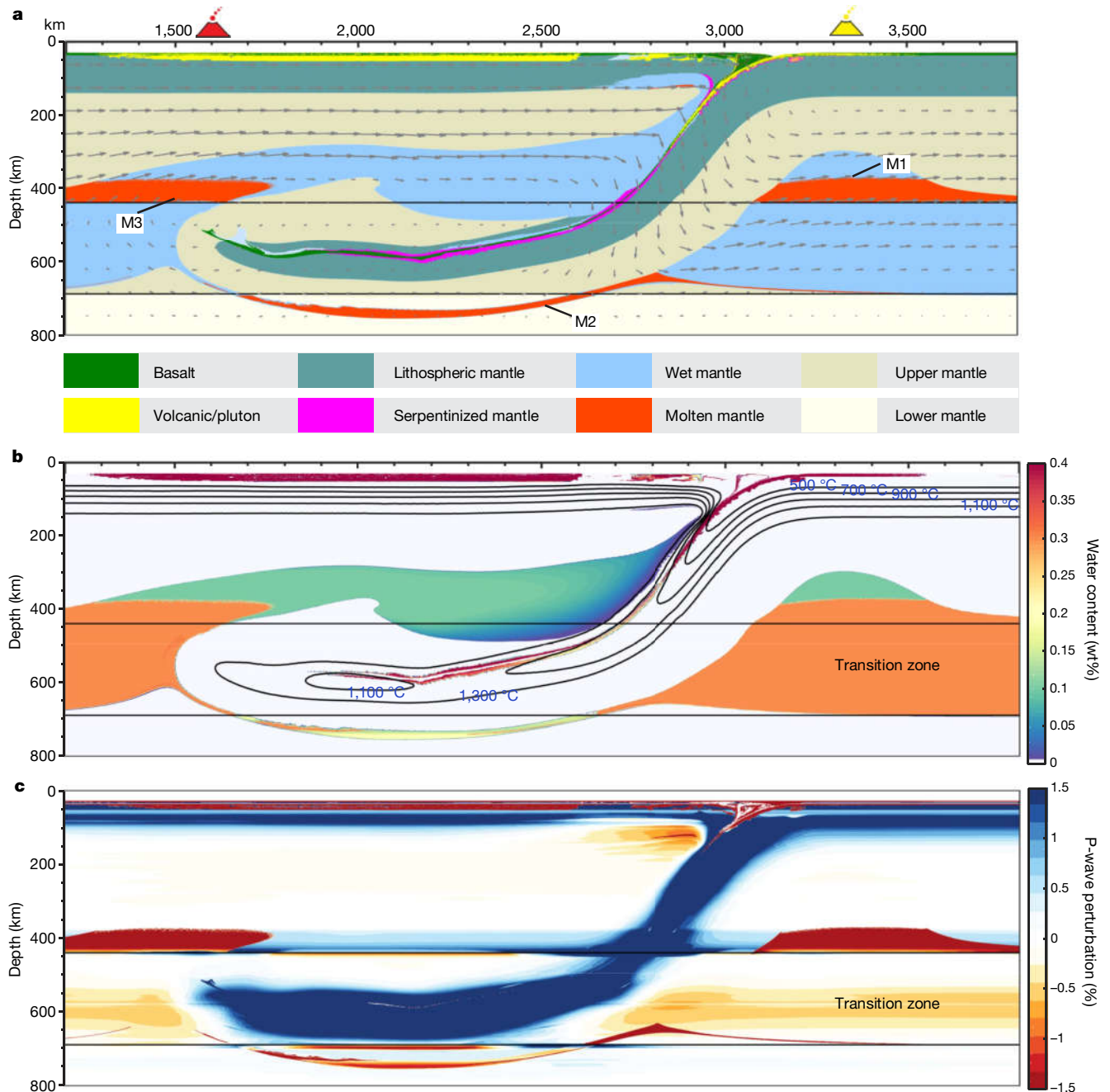


Fig. 2 | Dynamics of subduction-induced dehydration melting above and below the mantle transition zone. a, Composition field. A colour key indicating different rock types is given at the bottom. Two horizontal black lines mark depths of 410 km and 660 km. Three partially molten regions

(M1, M2 and M3) are indicated. **b**, Water content with temperature contours. **c**, Seismic P-wave velocity anomalies. An initial water content of 0.3 wt% is assumed in the MTZ, and the reference melt extraction timescale $t_{\text{ref}} = 6$ kyr (see Methods).

comprehensively explain the LVZs and the intraplate and petit-spot volcanism. This model does not exclude the devolatilization of the stagnant Pacific slab as a mechanism to explain the LVZ3 region and the overlying intraplate volcanism¹⁵, which favours the upwelling of volatile-rich plumes from the MTZ²⁹ as envisaged by the Big Mantle Wedge model^{15,30}. However, the same slab-derived volatiles cannot obviously be the cause of both the LVZ1 and LVZ2 and of the petit-spots, implying the presence of a metasomatized MTZ before the last subduction episode. The accumulation of water in the MTZ could be caused by, for example, delamination of volatile-rich lithospheric roots³¹, or by previous slab dehydration episodes in the MTZ and subsequent absorption of the water by wadsleyite and ringwoodite. Alongside with water, reduced (by redox-freezing) carbonated sources and restitic K-hollandite-bearing sediments are

required to explain the volatile-rich, alkaline and EM1-type petrological and geochemical signature of the basalts^{32,33}. This is not surprising, as the MTZ, a graveyard for stagnating slabs, is the most likely candidate to host volatiles and subducted sediments, and long-term isolation of these MTZ domains would be consistent with the ancient metasomatizing episodes estimated for intraplate basalts^{32–34}. Subsequent subduction events would mobilize the wet and (carbon + alkali)-bearing MTZ rocks, promoting the formation of silica-undersaturated magmas in the upper mantle. It is important to note that the addition of these components is not critical to our results, because the location and amounts of partial melting above and below the MTZ will still be dictated by the distribution of wet MTZ domains, while reduced carbonated sources are expected to experience redox melting at shallower depths (<250 km)³⁵.

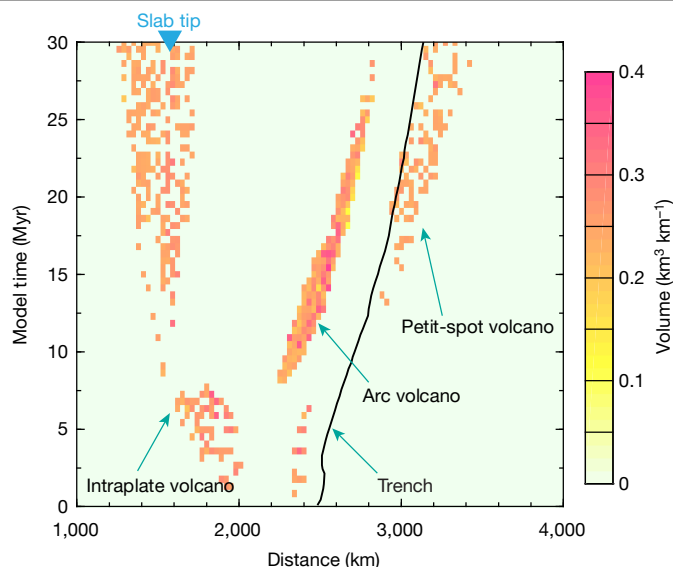


Fig. 3 | Volume of volcanics versus time. The volcanics include arc volcanoes produced by shallow decompression/hydrous melting and intraplate/petit-spot volcanoes produced by wet deep upper-mantle melting. The trench location (black line) through the model evolution and the final location of the slab tip (inverted blue triangle) are also indicated.

The process proposed here could potentially explain also the Cenozoic anorogenic volcanism in the Mediterranean¹³ and intraplate volcanism in the Turkish–Iranian Plateau³⁶ regions characterized by the long-term subduction of the Tethys Ocean. Together with surface intraplate/petit-spot volcanism, constraints on deep seismic low velocities and/or high electrical conductivity may thus indicate a volatile-rich and/or partially molten mantle within and around the transition zone.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2045-y>.

- Chen, Y., Zhang, Y., Graham, D., Su, S. & Deng, J. Geochemistry of Cenozoic basalts and mantle xenoliths in northeast China. *Lithos* **96**, 108–126 (2007).
- Wang, X.-C., Wilde, S. A., Li, Q.-L. & Yang, Y.-N. Continental flood basalts derived from the hydrous mantle transition zone. *Nat. Commun.* **6**, 7700 (2015).
- Chen, C. et al. Mantle transition zone, stagnant slab and intraplate volcanism in northeast Asia. *Geophys. J. Int.* **209**, 68–85 (2017).
- Hirano, N. et al. Volcanism in response to plate flexure. *Science* **313**, 1426–1428 (2006).
- Okumura, S. & Hirano, N. Carbon dioxide emission to Earth's surface by deep-sea volcanism. *Geology* **41**, 1167–1170 (2013).
- Machida, S. et al. Petit-spot geology reveals melts in upper-most asthenosphere dragged by lithosphere. *Earth Planet. Sci. Lett.* **426**, 267–279 (2015).
- Pilet, S. et al. Pre-subduction metasomatic enrichment of the oceanic lithosphere induced by plate flexure. *Nat. Geosci.* **9**, 898–903 (2016).

- Li, C., Van der Hilst, R. D., Meltzer, A. S. & Engdahl, E. R. Subduction of the Indian lithosphere beneath the Tibetan Plateau and Burma. *Earth Planet. Sci. Lett.* **274**, 157–168 (2008).
- Tauzin, B., Debayle, E. & Wittlinger, G. Seismic evidence for a global low-velocity layer within the Earth's upper mantle. *Nat. Geosci.* **3**, 718–721 (2010).
- Fukao, Y. & Obayashi, M. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity. *J. Geophys. Res. Solid Earth* **118**, 5920–5938 (2013).
- Liu, Z., Park, J. & Karato, S.-i. Seismological detection of low velocity anomalies surrounding the mantle transition zone in Japan subduction zone. *Geophys. Res. Lett.* **43**, 2480–2487 (2016).
- Wei, S. S. & Shearer, P. M. A sporadic low-velocity layer atop the 410 km discontinuity beneath the Pacific Ocean. *J. Geophys. Res. Solid Earth* **122**, 5144–5159 (2017).
- Lustrino, M. & Wilson, M. The circum-Mediterranean anorogenic Cenozoic igneous province. *Earth Sci. Rev.* **81**, 1–65 (2007).
- Tang, Y. et al. Changbaishan volcanism in northeast China linked to subduction-induced mantle upwelling. *Nat. Geosci.* **7**, 470–475 (2014).
- Zhao, D., Tian, Y., Lei, J., Liu, L. & Zheng, S. Seismic image and origin of the Changbai intraplate volcano in East Asia: role of big mantle wedge above the stagnant Pacific slab. *Phys. Earth Planet. Inter.* **173**, 197–206 (2009).
- Karato, S.-i. Water distribution across the mantle transition zone and its implications for global material circulation. *Earth Planet. Sci. Lett.* **301**, 413–423 (2011).
- Kelbert, A., Schultz, A. & Egbert, G. Global electromagnetic induction constraints on transition-zone water content variations. *Nature* **460**, 1003–1006 (2009).
- Bercovici, D. & Karato, S.-i. Whole-mantle convection and the transition-zone water filter. *Nature* **425**, 39–44 (2003).
- Liu, Z., Park, J. & Karato, S.-i. Seismic evidence for water transport out of the mantle transition zone beneath the European Alps. *Earth Planet. Sci. Lett.* **482**, 93–104 (2018).
- Schmandt, B., Jacobsen, S. D., Becker, T. W., Liu, Z. & Dueker, K. G. Dehydration melting at the top of the lower mantle. *Science* **344**, 1265–1268 (2014).
- Hier-Majumder, S. & Tauzin, B. Pervasive upper mantle melting beneath the western US. *Earth Planet. Sci. Lett.* **463**, 25–35 (2017).
- Mao, Z. et al. Elasticity of hydrous wadsleyite to 12 GPa: implications for Earth's transition zone. *Geophys. Res. Lett.* **35**, <https://doi.org/10.1029/2008GL035618> (2008).
- Irifune, T. et al. Sound velocities of majorite garnet and the composition of the mantle transition region. *Nature* **451**, 814–817 (2008).
- Bezada, M., Faccenda, M. & Toomey, D. Representing anisotropic subduction zones with isotropic velocity models: a characterization of the problem and some steps on a possible path forward. *Geochim. Geophys. Geosyst.* **17**, 3164–3189 (2016).
- Obayashi, M., Sugioka, H., Yoshimitsu, J. & Fukao, Y. High temperature anomalies oceanward of subducting slabs at the 410-km discontinuity. *Earth Planet. Sci. Lett.* **243**, 149–158 (2006).
- Zhao, D. & Tian, Y. Changbai intraplate volcanism and deep earthquakes in East Asia: a possible link? *Geophys. J. Int.* **195**, 706–724 (2013).
- Cline, C. J. II, Faul, U. H., David, E. C., Berry, A. J. & Jackson, I. Redox-influenced seismic properties of upper-mantle olivine. *Nature* **555**, 355–358 (2018).
- Xu, W., Lithgow-Bertelloni, C., Stixrude, L. & Ritsema, J. The effect of bulk composition and temperature on mantle seismic structure. *Earth Planet. Sci. Lett.* **275**, 70–79 (2008).
- Litasov, K. D., Shatskiy, A., Ohtani, E. & Yaxley, G. M. Solidus of alkaline carbonatite in the deep mantle. *Geology* **41**, 79–82 (2013).
- Kuritani, T. et al. Buoyant hydrous mantle plume from the mantle transition zone. *Sci. Rep.* **9**, 6549 (2019).
- Green, H. W., II, Chen, W.-P. & Brudzinski, M. R. Seismic evidence of negligible water carried below 400-km depth in subducting lithosphere. *Nature* **467**, 828–831 (2010).
- Mazza, S. E. et al. Sampling the volatile-rich transition zone beneath Bermuda. *Nature* **569**, 398–403 (2019).
- Wang, X.-J. et al. Mantle transition zone-derived EM1 component beneath NE China: geochemical evidence from Cenozoic potassic basalts. *Earth Planet. Sci. Lett.* **465**, 16–28 (2017).
- Kuritani, T., Ohtani, E. & Kimura, J. I. Intensive hydration of the mantle transition zone beneath China caused by ancient slab stagnation. *Nat. Geosci.* **4**, 713–716 (2011).
- Rohrbach, A. & Schmidt, M. W. Redox freezing and melting in the Earth's deep mantle resulting from carbon–iron redox coupling. *Nature* **472**, 209–212 (2011).
- Soltanmohammadi, A. et al. Transport of volatile-rich melt from the mantle transition zone via compaction pockets: implications for mantle metasomatism and the origin of alkaline lavas in the Turkish–Iranian plateau. *J. Petrol.* **59**, 2273–2310 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Modelling approach

The 2D petrological–thermomechanical numerical code I2VIS used in this study is based on a finite difference method using a marker-in-cell technique on a staggered grid³⁷. It solves mass, momentum and energy conservation equations (1)–(3) on the Eulerian grid and interpolates physical properties to the markers for advection accordingly.

$$\frac{\partial v_i}{\partial x_i} = 0 \quad (1)$$

$$-\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j} \left(\eta \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \right) + \rho g_i = 0 \quad (2)$$

$$\rho c_p \left(\frac{\partial T}{\partial t} + v_i \frac{\partial T}{\partial x_i} \right) = \frac{\partial}{\partial x_i} \left(k \frac{\partial T}{\partial x_i} \right) + H_r + H_s + H_a \quad (3)$$

where v_i is velocity, x_i coordinate, P dynamic pressure, ρ density, g gravity acceleration, c_p heat capacity, T temperature, k thermal conductivity, H_r radioactive heating, $H_s = \tau_{ij} \dot{\epsilon}_{ij}$ shear heating, and $H_a = T \alpha \frac{DP}{Dt}$ adiabatic heating, where $\frac{D}{Dt}$ is the material time derivative. The latent heat is implicitly considered by computing the effective thermal expansion and heat capacity.

Model configuration

The initial model set-up (6,000 × 1,000 km discretized with 1,501 × 501 nodes) is composed of a 3,500-km subducting plate and a 2,500-km overriding plate. The model imposes free-slip mechanical boundary condition at the top with 30-km-thick and viscosity of 10¹⁸ Pa s ‘sticky-air’ to mimic free surface; the bottom boundary is no slip, and side boundaries are periodic. The bottom no-slip condition is needed to define an initial horizontal velocity from which finite differences can be computed for this variable. Comparison with results from a model with a bottom free-slip condition and closed vertical walls indicate that the bottom no-slip boundary condition does not affect the subduction dynamics at all, as it is confined above the lower mantle. The initial thermal structure is defined by the half-space cooling age for the plates (50 Myr old) and an adiabatic thermal gradient of 0.5 K km^{−1} for the underlying mantle. The thermal boundary conditions are isothermal on the top and bottom, while side boundaries are periodic, consistent with the mechanical boundary conditions.

To initiate subduction, the subducting slab extends down to about 200 km in the upper mantle together with a rheologically weak zone on top of it which lubricate the initial contact between the plates. The high numerical resolution (4 km × 2 km) used here is needed to ensure plate contact lubrication at shallow depths and localized, bending-related hydration at the trench outer-rise. Tests at a lower resolution (4 km × 4 km) result in less-localized slab mantle hydration, whereas with a resolution of 8 km × 4 km, self-sustained subduction and slab rollback do not appear spontaneously.

Viscous–plastic rheological model

The rock mechanical behaviour is represented by the effective viscosity, which combines ductile (dislocation, diffusion and Peierls creep) and brittle (Drucker–Prager) deformation. The effective ductile viscosity is given by the harmonic average of the combined rheologies (parameters and physical meaning are defined in Extended Data Table 1):

$$\eta_{\text{ductile}} = \left(\frac{1}{\eta_{\text{disl}}} + \frac{1}{\eta_{\text{diff}}} + \frac{1}{\eta_{\text{Peierls}}} \right)^{-1} \quad (4)$$

where the dislocation and diffusion creep are given by³⁸:

$$\dot{\epsilon} = A(\sigma/\mu)^n (b/d)^m \exp\left(-\frac{E+PV}{RT}\right) \exp(\alpha\phi) \quad (5)$$

$$\eta = \frac{\sigma}{2\dot{\epsilon}} \quad (6)$$

For hydrated (wet) mantle, viscosity is reduced by $\eta_{\text{wet}} = \eta_{\text{dry}} \left(\frac{C_w}{C_{w0}} \right)^{-r/n}$, and C_w , C_{w0} are water content and reference water content (100 ppm, which is the water content for the dry upper mantle), respectively.

The Peierls creep η_{Peierls} is given by³⁹:

$$\eta_{\text{Peierls}} = 0.5A_p \sigma'^{-1} \exp\left\{ \frac{E_{\text{Peierls}} + PV_{\text{Peierls}}}{RT} \left[1 - \left(\frac{\sigma'_{\text{II}}}{\sigma_{\text{Peierls}}} \right)^p \right]^q \right\} \quad (7)$$

Parameters are defined in Extended Data Table 1.

Brittle behaviour occurs when stresses are above the plastic yield stress τ_y :

$$\eta_{\text{ductile}} \leq \frac{\tau_y}{2\dot{\epsilon}_{\text{II}}} \quad (8)$$

$$\tau_y = C + \mu P \quad (9)$$

Petrological modelling

Petrological solid–solid phase changes are included through the density and enthalpy look-up tables for basalt and pyrolite obtained from PERPLE_X⁴⁰. Therefore, phase transition boundaries at 410 km and 660 km have been considered.

The solidus ($T_s = f(P, T, H_2O)$) and liquidus ($T_l = f(P, T)$) temperatures for the upper mantle and MTZ are taken from high-pressure experiments⁴¹ (Extended Data Fig. 1). At lower mantle conditions, T_s and T_l vary considerably among different experiments. Here we adopt the dry solidus and liquidus of chondritic mantle⁴², as these are more compatible with the results of KLB-1 peridotite⁴³, while the wet solidus⁴⁴ was measured on samples with an estimated water content of 400 ppm wt.

A conservative estimate of the melt fraction in the wet upper mantle¹⁸ is applied:

$$\phi = \frac{W_a - W_{ol}}{W_m - W_{ol}} \quad (10)$$

where W_a , W_{ol} and W_m (10 wt%) are water mass fraction of the ambient mantle, olivine and melt, respectively. Note that the water solubility in olivine increases with pressure, so the melt fraction will decrease with depth if W_a remains constant (Extended Data Fig. 3).

The silicate melt density (Extended Data Fig. 2) is taken from high-pressure sink–float experiments⁴⁵, which show that the melt becomes denser than the surrounding mantle at around 400 km (refs. ^{45,46}) owing to the increased compressibility. However, the presence of water generally reduces the melt density such that it becomes buoyant relative to solid mantle (see Extended Data Fig. 2), rendering melt extraction at this depth possible. We also test the melt density from molecular dynamics simulations at high-pressure conditions⁴⁷ (Extended Data Figs. 4f, 5d).

Melt extraction timescale

The distance over which the compaction rate decreases by a factor of e is the characteristic length scale of the compaction process and is known as the compaction length, δ_c :

$$\delta_c = \sqrt{\frac{K \left(\zeta + \frac{4}{3} \eta \right)}{\eta_f}} \quad (11)$$

where ζ and η are the effective bulk and shear viscosities, respectively, of the partially molten rock; η_f is the fluid viscosity; K is the permeability given by the empirical equation:

$$K = K_0 \left(\frac{\phi}{\phi_0} \right)^n \quad (12)$$

where ϕ is the porosity (melt fraction); K_0 (10^{-12} m^2) is the permeability at the reference porosity ϕ_0 (0.01); and $n = 3$.

The relative migration velocity between the melt and the solid matrix is w :

$$w = \frac{K \Delta \rho g}{\eta_f \phi} \quad (13)$$

Thus, the extraction timescale t :

$$t = \frac{\delta_c}{w} = \frac{\phi}{\Delta \rho g} \sqrt{\frac{\eta_f \left(\zeta + \frac{4}{3} \eta \right)}{K}} \quad (14)$$

where $\zeta \approx \frac{\eta}{\phi}$, so $\zeta + \frac{4}{3} \eta \approx \frac{\eta}{\phi}$; if $\eta_f = 1 \text{ Pa s}$, then

$$t \approx \frac{1,000}{\Delta \rho g \phi} \sqrt{\eta} \quad (15)$$

where $\Delta \rho$ is the density difference between the solid and melt, ranging from about -100 kg m^{-3} to 270 kg m^{-3} and typically about $70\text{--}180 \text{ kg m}^{-3}$ in the lowermost upper mantle if most of the water is partitioned into melt, and the surrounding mantle viscosity $\eta = 10^{19}\text{--}10^{20} \text{ Pa s}$ in the upper mantle. For melt fraction $\phi = 0.02$, the estimated timescale would be $t_{\text{ref}} \approx 3\text{--}20 \text{ kyr}$. Note that this timescale is not the time for melt migration to the surface, but only illustrates the efficiency of melt segregation from the solid matrix and the likelihood of its emplacement at shallow depths. Indeed, a small migration time implies a large $\Delta \rho$ (that is, high buoyancy force) and/or high melt fraction (that is, high permeability) and/or weak solid matrix which can be easily deformed during compaction/decompaction processes. In this study, we use a reference value $t_{\text{ref}} = 6 \text{ kyr}$.

Melt extraction depends not only on these three parameters (and fluid viscosity), but also on the dihedral angle (that is, melt interconnectivity). Previous experiments showed that the dihedral angle decreases systematically with increasing pressure such that it probably allows for complete wetting at about 400 km (ref. ⁴⁸) where the dihedral angle is $< 5^\circ$ (ref. ⁴⁹). At these depths, melt interconnectivity is high even for low amounts of melt ($< 1\%$), which makes melt extraction possible provided that the extraction timescale is sufficiently low (that is, the melt migration process is efficient and not hindered by the solid matrix). As such, when the extraction timescale t is smaller than t_{ref} , the material is extracted at the surface forming plutonic intrusions or volcanics⁵⁰. Note that when the density of the melt is higher than that of the solid surrounding mantle, as occurs between 11.5 GPa and 13.5 GPa when the water content is low⁴⁵ (Extended Data Fig. 2b), there will be no melt extraction. In these conditions, the denser melt should percolate downward and accumulate over the 410-km discontinuity¹⁸. However, dry melting generally does not occur at ambient mantle conditions, except when there is an abnormal heat source associated with a mantle plume. As hydrogen partitions preferentially into the melt, the water content in the melt would be quite high, decreasing its density^{45,46,49}. As a result, hydrous melt should be less dense than the solid matrix throughout the upper mantle⁴⁵.

The melt migration process is illustrated here with more realistic models accounting for visco-elastoplastic deformation in a two-phase flow regime. These models demonstrate that melt migration from the deep upper mantle to the surface should occur through several mechanisms: viscous diapirism, viscoplastic decompaction channels and elastoplastic dyking^{51,52} (Extended Data Fig. 6). For weak host rocks

where viscous deformation dominates, such as the asthenosphere, magma migrates by diapirism. When the magma moves through the lithosphere–asthenosphere boundary (or the lower crust in continents) where both ductile and brittle deformation occur, the fluid compaction pressure might reach the tensile strength, and magma could migrate by channelling. If the host rock is completely elastoplastic, such as the core of lithospheric mantle and upper crust, magma migrates by dyking.

Water budget

The phase diagram reporting the maximum water content that can be hosted in hydrated or wet mantle rocks (that is, absorbed by nominally anhydrous minerals, NAMs) is built upon the compilation from refs. ^{41,53} (Extended Data Fig. 3). It is often assumed that heterogeneously serpentinized mantle rocks below the oceanic Moho can contain up to $2 \text{ wt\% H}_2\text{O}$ (refs. ^{54,55}). As such, the maximum water content in hydrated rocks⁵³ is scaled accordingly.

As the oceanic crust completely dehydrates at about 300 km depth⁴¹, generating fluids that fuel arc-volcanism only, we assume a dry crust for the sake of simplicity. On the other hand, dehydration of the underlying mantle within the transition zone is thought to cause intracontinental magmatism¹⁵. Consequently, we allow for serpentinization by bending-related deformation when the strain of mantle rocks is greater than 0.1 (ref. ⁵⁵).

When the rock water content exceeds the saturation limit, decomposition of hydrous minerals or water exsolution in NAMs occurs, and fluid markers are generated and migrate according to Darcy's law^{55,56} until they are absorbed by dry markers:

$$V_i^f = V_i^s - \frac{\nabla P - \rho_f \mathbf{g}_i}{(\rho_s - \rho_f) g_y} V_0 \quad (16)$$

where V_i^s, V_i^f are the velocities of solid and fluid phases, respectively; ρ_s, ρ_f are the densities of solid and fluid, respectively; V_0 is a constant percolation velocity; \mathbf{g} is the gravity acceleration vector as defined in equation (2), and g_y is its vertical component.

Upon partial melting and extraction, the water is partitioned into the extracted melt and water in the residual peridotite as:

$$C_w^{\text{melt}} = \frac{C_w}{\phi(1-D) + D} \quad (17)$$

$$C_w^{\text{res}} = \frac{C_w - \phi C_w^{\text{melt}}}{(1-\phi)} \quad (18)$$

where $D = 0.01$ is the hydrogen partition coefficient for olivine polymorphs.

Falling block tests. The validity of the petrological model used here can be easily tested with a simple model in which a falling block (simulating the subducting slab) sinks into the wet MTZ, exciting wet upwellings to the upper mantle and squeezing water into the lower mantle (Extended Data Fig. 4). These tests indicate that the melt layer gets thicker ($> 100 \text{ km}$) when melt extraction is not efficient, owing to very small amounts of melt/water and/or denser melt phase. This might explain the thick low-velocity layers above 410 km in many regions⁹. After melt extraction, less water remains above the transition zone, causing higher viscosity and less melt fraction, which yields a larger extraction timescale: that is, the melt preferentially ponds above 410 km depth.

Seismic velocity anomalies. The seismic velocity perturbation in Fig. 2c have been computed as:

$$\delta \ln V = \frac{V - V_{\text{ref}}}{V_{\text{ref}}} \quad (19)$$

where V_{ref} is the average seismic velocity at specific depth.

The change of seismic wave velocities caused by the existence of a fluid phase is given by:

$$\frac{V_s}{V_s^0} = \frac{\sqrt{N/\mu}}{\sqrt{\bar{\rho}/\rho}} \quad (20)$$

and

$$\frac{V_p}{V_p^0} = \frac{\sqrt{\frac{K_{\text{eff}}}{k} + (4\gamma/3)N/\mu}}{\sqrt{1 + 4\gamma/3 \cdot \sqrt{\bar{\rho}/\rho}}} \quad (21)$$

where

$$\frac{K_{\text{eff}}}{k} = \frac{K_b}{k} + \frac{(1 - K_b/k)^2}{1 - \phi - (K_b/k) + (\phi k/k_f)} \quad (22)$$

and $\gamma = \frac{\mu}{k} = \frac{3(1-2\nu)}{2(1+\nu)} \cdot V_s^0/V_p^0$ are the shear and compressional wave velocities of the solid phase; k , μ , ν and ρ are the bulk modulus, shear modulus, Poisson's ratio and density of the solid phase, respectively. $\bar{\rho} = (1 - \phi)\rho + \phi\rho_f$ is the effective density when fluid (for example, melt) exists. K_b and N are the bulk and shear moduli, which are dependent on melt fraction and dihedral angle⁵⁷:

$$K_b = (1 - \phi)k(1 - (1 - \phi)^{n_k}) \quad (23)$$

$$N = (1 - \phi)\mu(1 - (1 - \phi)^{n_\mu}) \quad (24)$$

where

$$n_k = a_1\phi + a_2(1 - \phi) + a_3\phi(1 - \phi)^{1.5} \quad (25)$$

$$n_\mu = b_1\phi + b_2(1 - \phi) + b_3\phi(1 - \phi)^2 \quad (26)$$

and ϕ is the dihedral angle

$$\phi = \frac{2A_{ss}}{2A_{ss} + A_{sl}} \quad (27)$$

with A_{ss} , A_{sl} the area of solid–solid contact and solid–liquid contact, respectively⁵⁸.

Extended Data Fig. 7 shows K_b/k and N/μ for the equilibrium geometry model at various dihedral angles.

Data availability

The dataset generated during the current study is available at https://figshare.com/articles/Yang_Faccenda_Nature2019/9933056.

Code availability

Requests about the numerical modelling codes associated with this paper should be sent to the main code developer (taras.gerya@erdw.ethz.ch). The map in Fig. 1a is created with open software GMT 5.4.3 which is under a GNU Lesser General Public License. The numerical 2D finite element code MVEP2 (<https://bitbucket.org/bkaus/mvep2>) was used for the two-phase flow model in Extended Data Fig. 6.

37. Gerya, T. V. & Yuen, D. A. Characteristics-based marker-in-cell method with conservative finite-differences schemes for modeling geological flows with strongly variable transport properties. *Phys. Earth Planet. Inter.* **140**, 293–318 (2003).
38. Karato, S.-i. & Wu, P. Rheology of the upper mantle: a synthesis. *Science* **260**, 771–778 (1993).
39. Kameyama, M., Yuen, D. A. & Karato, S.-i. Thermal-mechanical effects of low-temperature plasticity (the Peierls mechanism) on the deformation of a viscoelastic shear zone. *Earth Planet. Sci. Lett.* **168**, 159–172 (1999).
40. Connolly, J. Computation of phase equilibria by linear programming: a tool for geodynamic modeling and its application to subduction zone decarbonation. *Earth Planet. Sci. Lett.* **236**, 524–541 (2005).

41. Litasov, K. Physicochemical conditions for melting in the Earth's mantle containing a C–O–H fluid (from experimental data). *Russ. Geol. Geophys.* **52**, 475–492 (2011).
42. Andraut, D. et al. Melting of subducted basalt at the core-mantle boundary. *Science* **344**, 892–895 (2014).
43. Zhang, J. & Herzberg, C. Melting experiments on anhydrous peridotite KLB-1 from 5.0 to 22.5 GPa. *J. Geophys. Res. Solid Earth* **99**, 17729–17742 (1994).
44. Nomura, R. et al. Low core–mantle boundary temperature inferred from the solidus of pyrolite. *Science* **343**, 522–525 (2014).
45. Sakamaki, T., Suzuki, A. & Ohtani, E. Stability of hydrous melt at the base of the Earth's upper mantle. *Nature* **439**, 192–194 (2006).
46. Jing, Z. & Karato, S.-i. Effect of H₂O on the density of silicate melts at high pressures: static experiments and the application of a modified hard-sphere model of equation of state. *Geochim. Cosmochim. Acta* **85**, 357–372 (2012).
47. Guillot, B. & Sator, N. A computer simulation study of natural silicate melts. Part II: High pressure properties. *Geochim. Cosmochim. Acta* **71**, 4538–4556 (2007).
48. Yoshino, T., Nishihara, Y. & Karato, S.-i. Complete wetting of olivine grain boundaries by a hydrous melt near the mantle transition zone. *Earth Planet. Sci. Lett.* **256**, 466–472 (2007).
49. Freitas, D. et al. Experimental evidence supporting a global melt layer at the base of the Earth's upper mantle. *Nat. Commun.* **8**, 2186 (2017).
50. Sizova, E., Gerya, T., Brown, M. & Perchuk, L. Subduction styles in the Precambrian: insight from numerical experiments. *Lithos* **116**, 209–229 (2010).
51. Keller, T., May, D. A. & Kaus, B. J. P. Numerical modelling of magma dynamics coupled to tectonic deformation of lithosphere and crust. *Geophys. J. Int.* **195**, 1406–1442 (2013).
52. Lehmann, R. *Modelling of Magma Dynamics from the Mantle to the Surface* (Universitätsbibliothek Mainz, 2016).
53. Iwamori, H. Phase relations of peridotites under H₂O-saturated conditions and ability of subducting plates for transportation of H₂O. *Earth Planet. Sci. Lett.* **227**, 57–71 (2004).
54. van Keken, P. E., Hacker, B. R., Syracuse, E. M. & Abers, G. A. Subduction factory: 4. Depth-dependent flux of H₂O from subducting slabs worldwide. *J. Geophys. Res. Solid Earth* **116**, <https://doi.org/10.1029/2010JB007922> (2011).
55. Faccenda, M., Gerya, T. V. & Burlini, L. Deep slab hydration induced by bending-related variations in tectonic pressure. *Nat. Geosci.* **2**, 790–793 (2009).
56. Faccenda, M., Gerya, T. V., Mancktelow, N. S. & Moresi, L. Fluid flow during slab unbending and dehydration: implications for intermediate-depth seismicity, slab weakening and deep water recycling. *Geochim. Geophys. Geosystems* **13**, Q01010 (2012).
57. Takei, Y. Effect of pore geometry on V_p/V_s from equilibrium geometry to crack. *J. Geophys. Res. Solid Earth* **107**, 2043 (2002).
58. von Bagen, N. & Waff, H. S. Permeabilities, interfacial areas and curvatures of partially molten systems: results of numerical computations of equilibrium microstructures. *J. Geophys. Res. Solid Earth* **91**, 9261–9276 (1986).
59. Litasov, K. D. & Ohtani, E. Phase relations in hydrous MORB at 18–28 GPa: implications for heterogeneity of the lower mantle. *Phys. Earth Planet. Inter.* **150**, 239–263 (2005).
60. Pradhan, G. K. et al. Melting of MORB at core–mantle boundary. *Earth Planet. Sci. Lett.* **431**, 247–255 (2015).
61. Andraut, D. et al. Solidus and liquidus profiles of chondritic mantle: implication for melting of the Earth across its history. *Earth Planet. Sci. Lett.* **304**, 251–259 (2011).
62. Andraut, D. et al. Deep and persistent melt layer in the Archaean mantle. *Nat. Geosci.* **11**, 139–143 (2018).
63. Fiquet, G. et al. Melting of peridotite to 140 gigapascals. *Science* **329**, 1516–1518 (2010).
64. Boukaré, C. E., Ricard, Y. & Fiquet, G. Thermodynamics of the MgO–FeO–SiO₂ system up to 140 GPa: application to the crystallization of Earth's magma ocean. *J. Geophys. Res. Solid Earth* **120**, 6085–6101 (2015).
65. Baron, M. A. et al. Experimental constraints on melting temperatures in the MgO–SiO₂ system at lower mantle pressures. *Earth Planet. Sci. Lett.* **472**, 186–196 (2017).
66. Walter, M. J. et al. The stability of hydrous silicates in Earth's lower mantle: experimental constraints from the systems MgO–SiO₂–H₂O and MgO–Al₂O₃–SiO₂–H₂O. *Chem. Geol.* **418**, 16–29 (2015).
67. Sanloup, C. et al. Structure and density of molten fayalite at high pressure. *Geochim. Cosmochim. Acta* **118**, 118–128 (2013).
68. Bajgain, S., Ghosh, D. B. & Karki, B. B. Structure and density of basaltic melts at mantle conditions from first-principles simulations. *Nat. Commun.* **6**, 8578 (2015).
69. Agee, C. B. Crystal-liquid density inversions in terrestrial and lunar magmas. *Phys. Earth Planet. Inter.* **107**, 63–74 (1998).
70. Petitgirard, S. et al. Fate of MgSiO₃ melts at core–mantle boundary conditions. *Proc. Natl Acad. Sci. USA* **112**, 14186–14190 (2015).

Acknowledgements T. Gerya provided the I2VIS code. We acknowledge discussions with A. Marzoli, C. Meyzen, P. Nimis, D. Novella, M. Lustrino, K. Litasov, S.-i. Karato and X. Xu. J.Y. was financially supported by Dipartimento di Geoscienze, Università di Padova. M.F. acknowledges the European Research Council Starting Grant 758199.

Author contributions M.F. conceived the study. J.Y. performed all the numerical experiments and wrote the first draft of the paper. Both authors contributed equally to the discussion of the results and to the conclusions of this study.

Competing interests The authors declare no competing interests.

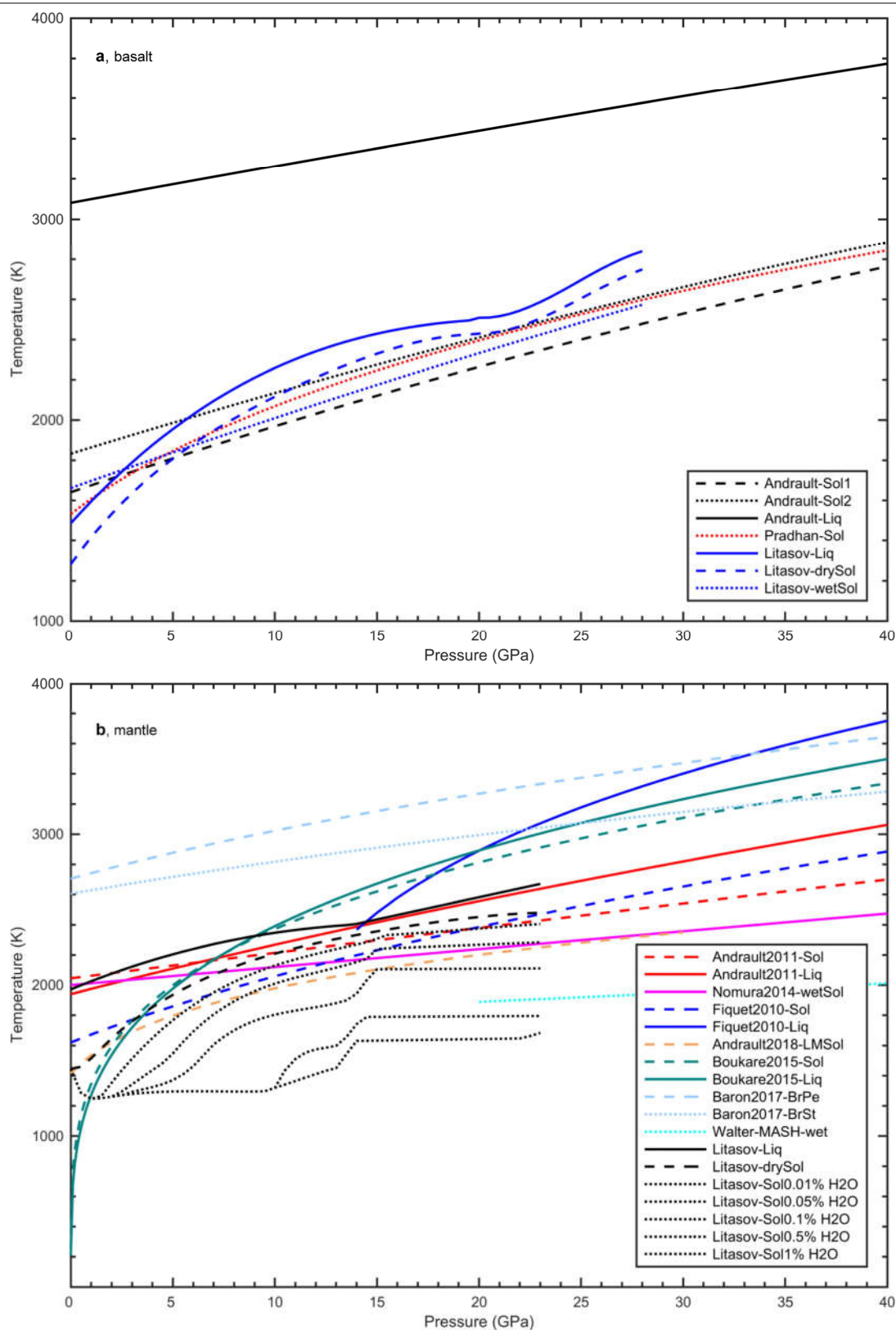
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2045-y>.

Correspondence and requests for materials should be addressed to M.F.

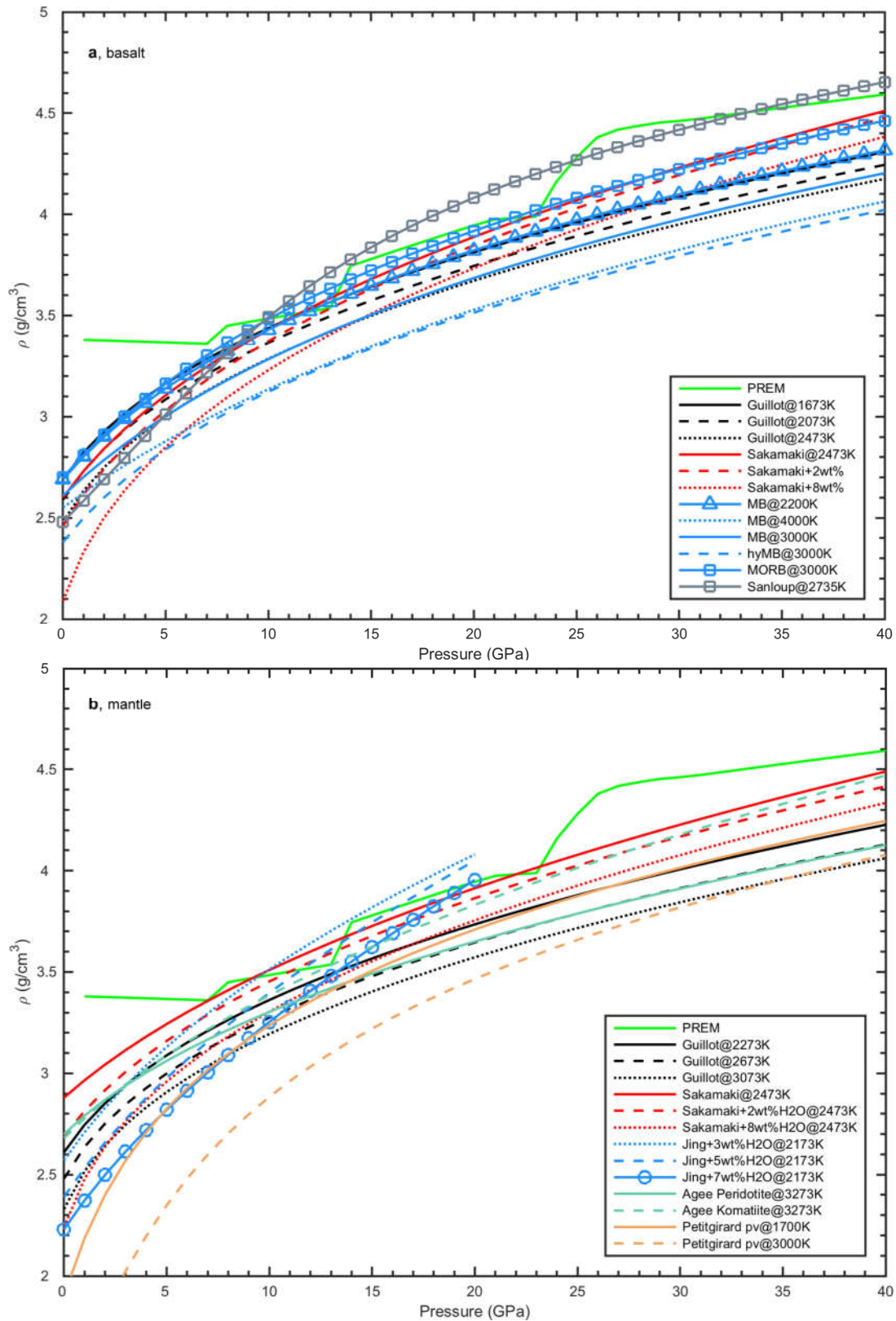
Peer review information Nature thanks Esteban Gazel, Dapeng Zhao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



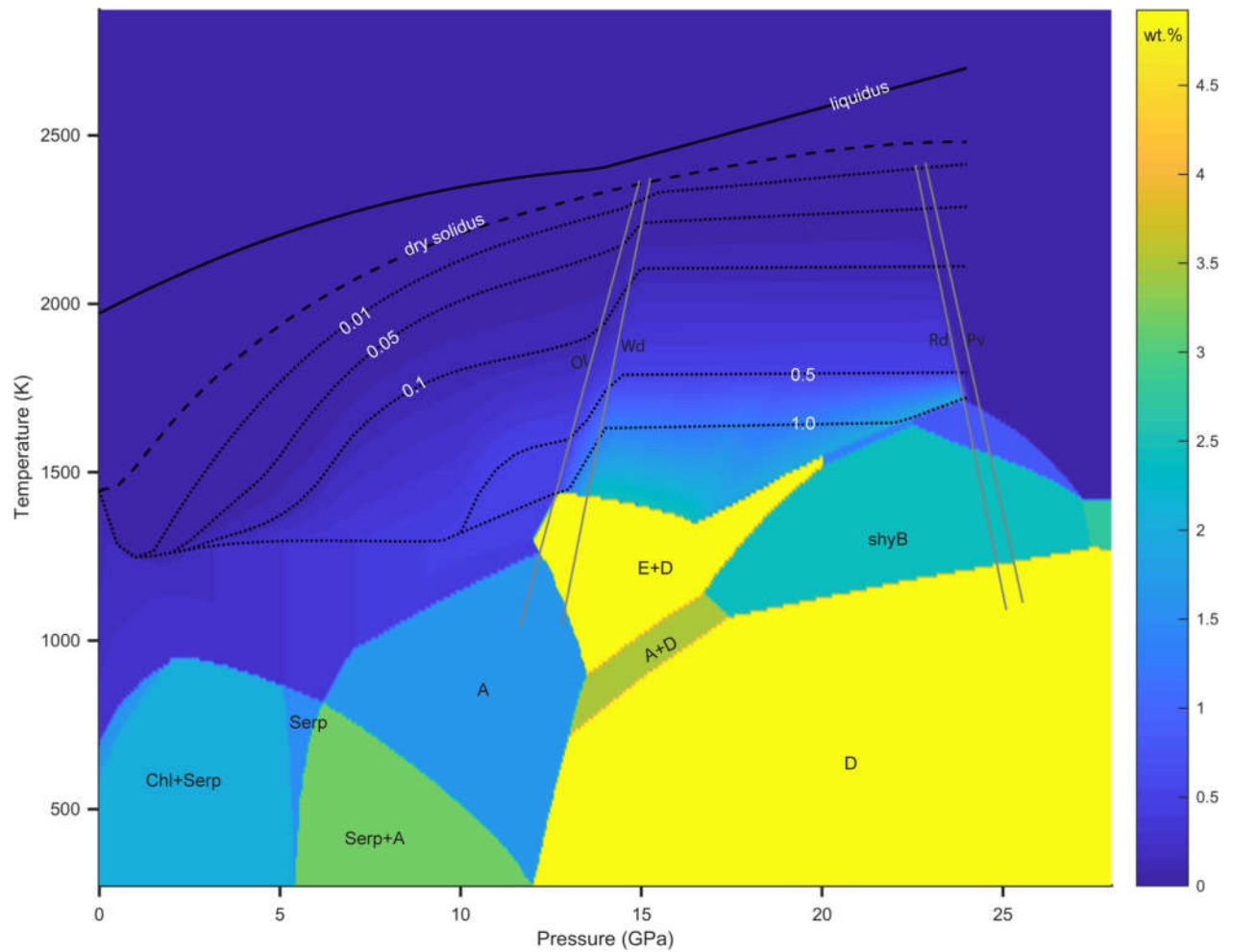
Extended Data Fig. 1|Solidus and liquidus of basalt and mantle. a, The solidus and liquidus of basalt are obtained from experimental data^{42,59,60}. The solidus from ref.⁶⁰ fits well within the uncertainty region of ref.⁴² and is thus adopted.

b, Solidus and/or liquidus of mantle collected from literature. Sol, solidus; Liq, liquidus; BrPe, MgSiO₃-MgO (bridgmanite + periclase); Fiquel, MgSiO₃-SiO₂ (bridgmanite + stishovite). Experimental data are from refs.^{41,44,61-66}.



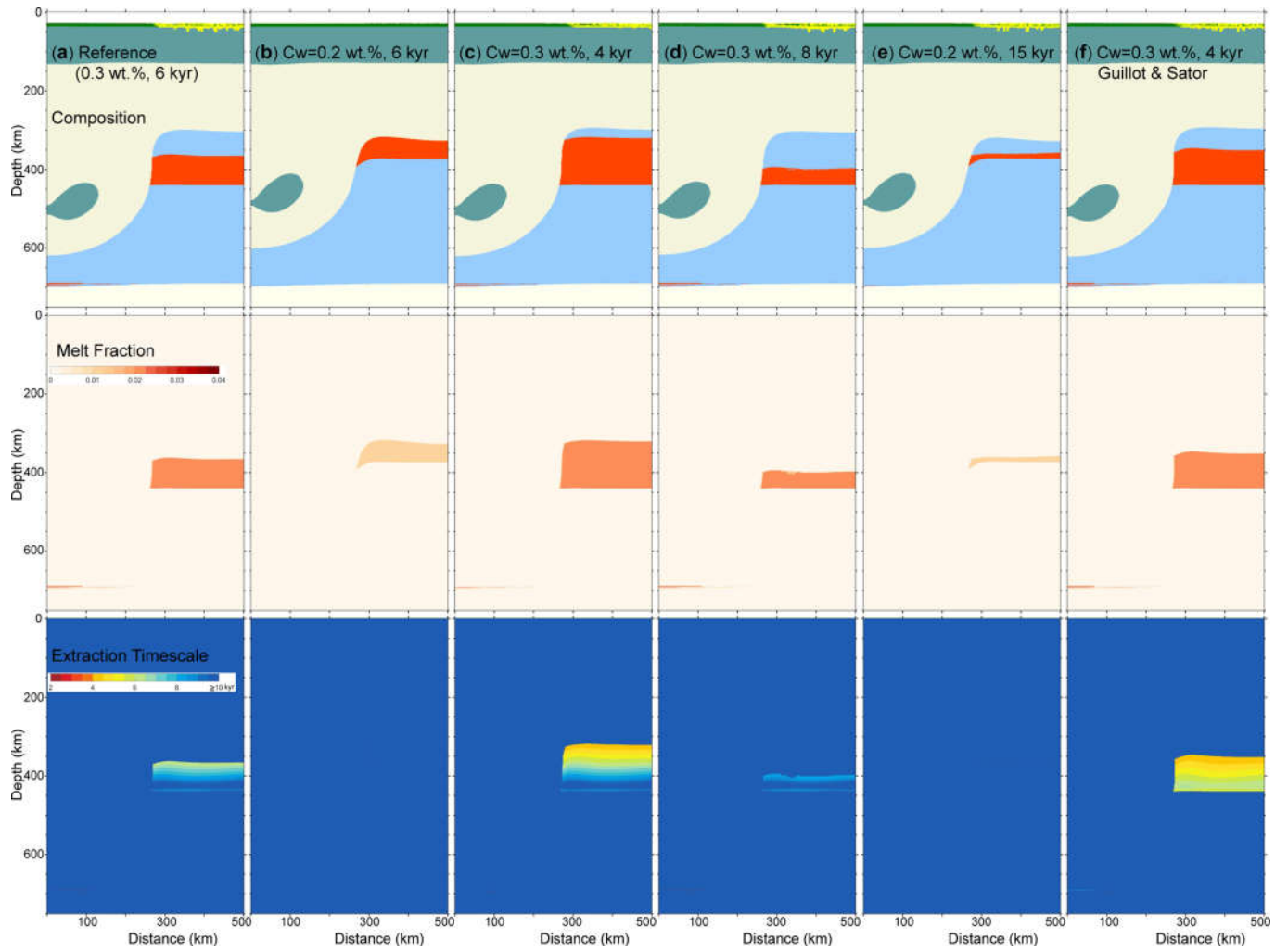
Extended Data Fig. 2 | Melt density of basalt and mantle for different temperatures and/or water contents. a, Basalt. PREM, density profile from Preliminary Reference Earth Model; dry melt density at temperatures of 1,673 K, 2,073 K and 2,473 K (ref. ⁴⁷) and 2,735 K (ref. ⁶⁷); dry and wet with 2 wt% and 8 wt% H₂O melt density at 2,473 K (ref. ⁴⁵); the modelled basalt (MB),

hydrated basalt (hyMB) and basalt (MORB)⁶⁸. **b, Mantle.** Melt density of dry peridotite⁴⁷; dry and wet (2 wt% and 8 wt% H₂O)⁴⁵; wet peridotite (3 wt%, 5 wt%, 7 wt% H₂O)⁴⁶; dry peridotite and komatiite⁶⁹ and perovskite (pv)⁷⁰. Note the density crossover at around 13 GPa (refs. ^{45,46}). All the profiles are fitted by third or fourth order of the Birch–Murnaghan equation of state.



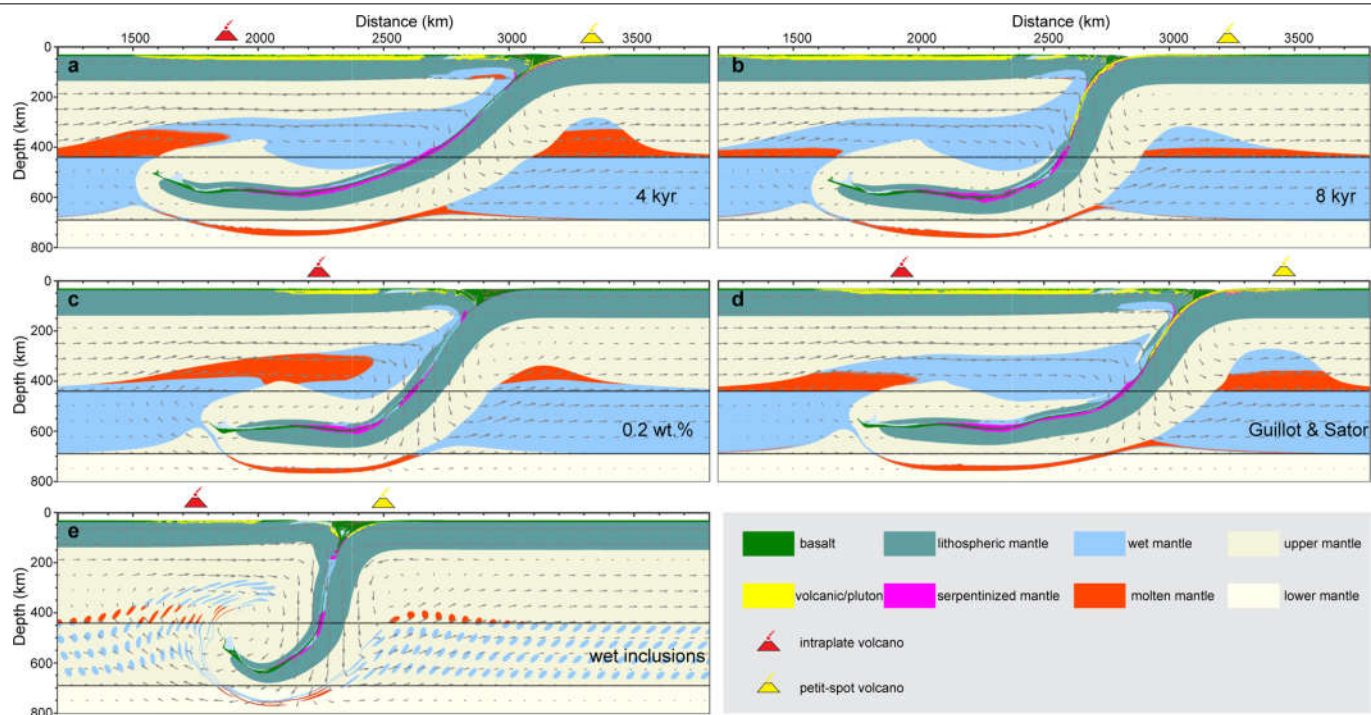
Extended Data Fig. 3 | Phase diagram of H₂O-peridotite, after ref.⁵³. The solidus/liquidus curves are the same as in Extended Data Fig. 1. The grey lines are olivine-wadsleyite (Ol-Wd) and ringwoodite-perovskite (Rd-Pv) phase

boundaries. The abbreviations of major hydrous phases are as follows: Chl, chlorite; Serp, serpentine; A, phase A; E, phase E; shyB, superhydrous phase B; D, phase D.



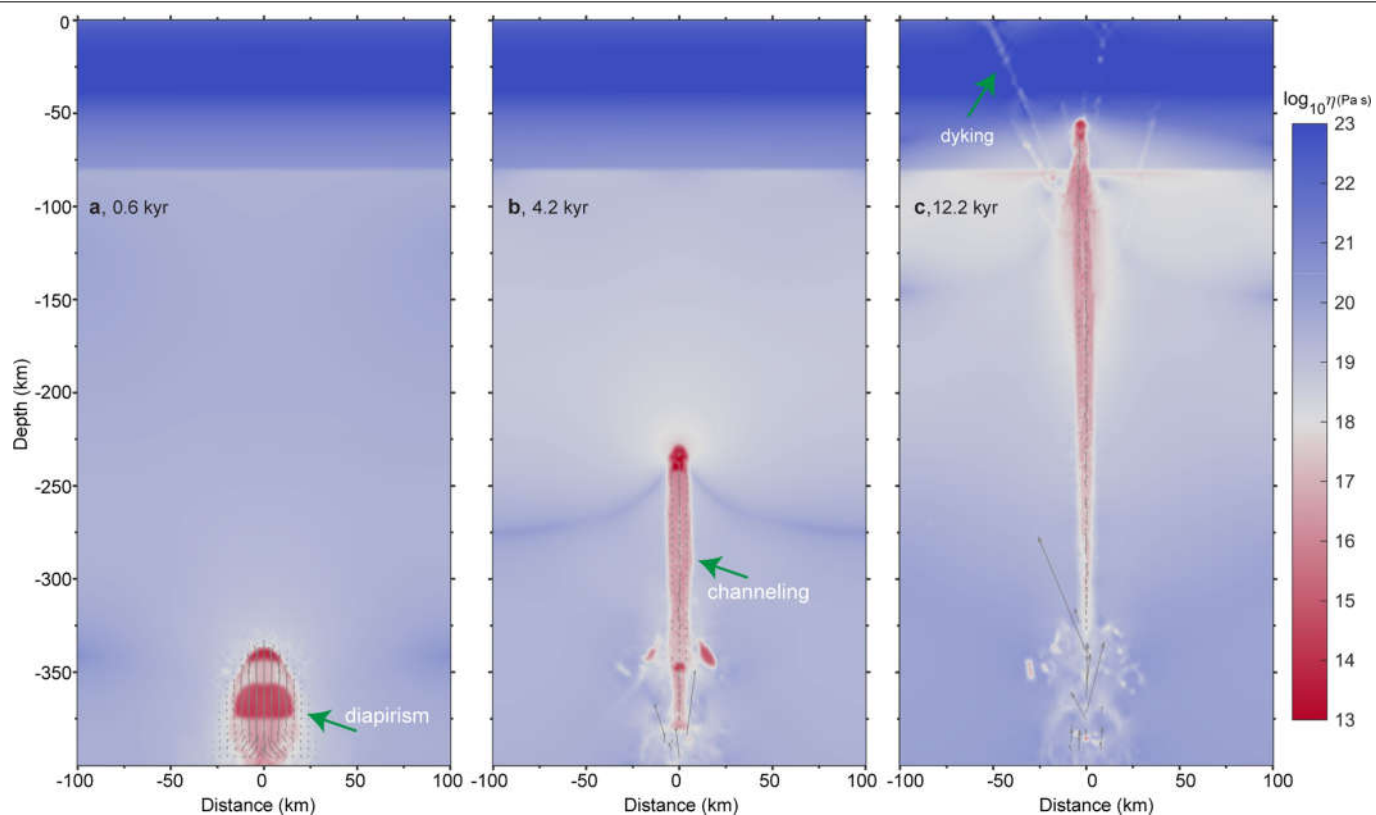
Extended Data Fig. 4 | Falling block simulations with different parameters. **a**, Reference model with initial MTZ water content of 0.3 wt%, melt density from ref. ⁴⁵ and reference extraction timescale $t_{\text{ref}} = 6$ kyr. **b–f**, Other tests are similar to this model except for **(b)** initial water content $C_w = 0.2$ wt%,

(c) extraction timescale $t_{\text{ref}} = 4$ kyr, **(d)** $t_{\text{ref}} = 8$ kyr, **(e)** $C_w = 0.2$ wt% and $t_{\text{ref}} = 15$ kyr, **(f)** $C_w = 0.3$ wt% and $t_{\text{ref}} = 4$ kyr by using the melt density from ref. ⁴⁷. Note that the extraction timescale is calculated only when the melt is less dense than the solid matrix.



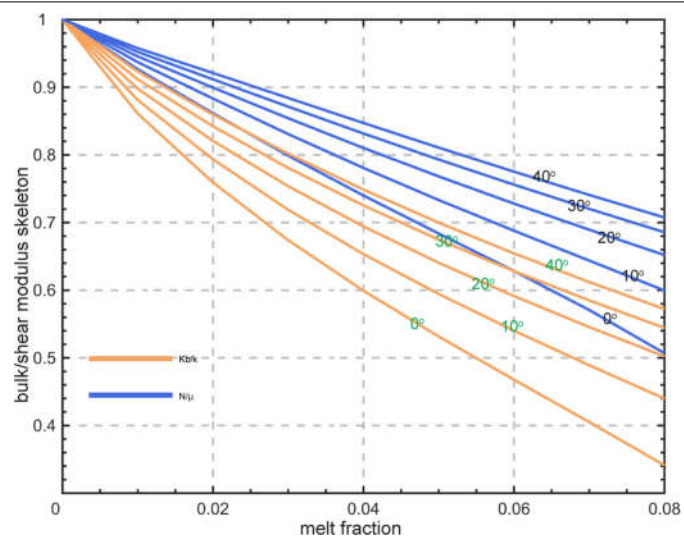
Extended Data Fig. 5 | Additional parameter tests. a, b, Extraction timescales of 4 kyr (**a**) and 8 kyr (**b**), with 0.3 wt% initial water content in both. **c,** Initial water content 0.2 wt%. **d,** Melt density from ref. ⁴⁷ and $t_{\text{ref}} = 4$ kyr. **e,** Wet

inclusions in the transition zone with $t_{\text{ref}} = 6$ kyr. Note that all the models differ by only one parameter from the reference model (Fig. 2), except **d**.



Extended Data Fig. 6 | Visco-plastic shear viscosity for melt percolation in two-phase flow. a–c, Melt percolation at three typical stages as (a), diapirism (b), channelling and (c), dyking from deep mantle to the surface. The numerical 2D finite element code MVEP2 was used to simulate melt migration dynamics. A small background strain rate (10^{-15} s^{-1} ; the model domain was extended by only

0.75 km after 12.2 kyr) was applied at the side boundaries. The top boundary is free surface. An initial porosity at the bottom boundary with Gaussian distribution (resulting in an average porosity of 0.127) was applied. The details of the approach allowing for its reproduction are provided elsewhere^{51,52}.



Extended Data Fig. 7 | Normalized bulk modulus K_b/k and shear modulus N/μ of skeleton (solid porous matrix) versus melt fraction. The ratios of both bulk and shear modulus decrease with melt fraction. The numbers shown on the lines are dihedral angles.

Extended Data Table 1 | Physical properties of rocks used in this study

Property	Symbol	Unit	Value
Gravity	g	m/s ²	9.81
Water content	C _w	wt. %	-
Reference water content	C _{w0}	wt. %	0.01
Melt fraction	ϕ	-	-
Melt-weakening factor	α	-	28
Shear modulus	μ	GPa	80
<i>Diffusion creep</i>			
prefactor	A	s ⁻¹	8.7×10 ¹⁵
Activation energy	E	kJ mol ⁻¹	300
Activation volume	V	cm ³ mol ⁻¹	6
Burgers vector	b	nm	0.5
Grain-size exponent	m	-	2.5
Water exponent	r	-	0.8
<i>Dislocation creep</i>			
prefactor	A	s ⁻¹	3.5×10 ²²
Activation energy	E	kJ mol ⁻¹	540
Activation volume	V	cm ³ mol ⁻¹	20
Stress exponent	n	-	3.5
Water exponent	r	-	1.2
<i>Peierls creep</i>			
prefactor	A _p	Pa ² s	10 ^{4.2}
Activation energy	E _{Peierls}	kJ mol ⁻¹	532
Activation volume	V _{Peierls}	cm ³ mol ⁻¹	12
Peierls stress	σ _{Peierls}	GPa	9.1
Exponent	p, q	-, -	1, 2
<i>Yield stress τ_y</i>			
Cohesion	C	MPa	10
Friction coefficient	μ	-	0.6

A simple dynamic model explains the diversity of island birds worldwide

<https://doi.org/10.1038/s41586-020-2022-5>

Received: 19 March 2019

Accepted: 22 January 2020

Published online: 19 February 2020

 Check for updates

Luis Valente^{1,2,3,4}✉, Albert B. Phillimore⁵, Martim Melo^{6,7,8}, Ben H. Warren⁹, Sonya M. Clegg^{10,11}, Katja Havenstein⁴, Ralph Tiedemann⁴, Juan Carlos Illera¹², Christophe Thébaud¹³, Tina Aschenbach¹ & Rampal S. Etienne³

Colonization, speciation and extinction are dynamic processes that influence global patterns of species richness^{1–6}. Island biogeography theory predicts that the contribution of these processes to the accumulation of species diversity depends on the area and isolation of the island^{7,8}. Notably, there has been no robust global test of this prediction for islands where speciation cannot be ignored⁹, because neither the appropriate data nor the analytical tools have been available. Here we address both deficiencies to reveal, for island birds, the empirical shape of the general relationships that determine how colonization, extinction and speciation rates co-vary with the area and isolation of islands. We compiled a global molecular phylogenetic dataset of birds on islands, based on the terrestrial avifaunas of 41 oceanic archipelagos worldwide (including 596 avian taxa), and applied a new analysis method to estimate the sensitivity of island-specific rates of colonization, speciation and extinction to island features (area and isolation). Our model predicts—with high explanatory power—several global relationships. We found a decline in colonization with isolation, a decline in extinction with area and an increase in speciation with area and isolation. Combining the theoretical foundations of island biogeography^{7,8} with the temporal information contained in molecular phylogenies¹⁰ proves a powerful approach to reveal the fundamental relationships that govern variation in biodiversity across the planet.

A key feature of global diversity is the tendency for some areas to harbour many more species than others^{7,8}. Uncovering the drivers and regulators of spatial differences in diversity of simple systems such as islands is a crucial step to understanding the global distribution of species richness. The two most prominent biodiversity patterns in fragmented or isolated environments worldwide are the increase in species richness with area and the decline in species richness with isolation^{8,11–14}. In their theory of island biogeography, MacArthur and Wilson proposed how the processes of colonization and extinction could explain these patterns^{7,8}. They argued that the rates of these processes are determined by the geographical context: colonization decreases with isolation and extinction decreases with area^{7,8}. They also suggested that rates of formation of island endemic species through in situ speciation increase with island isolation and area⁸. Despite an abundance of studies over five decades that support the general patterns predicted by MacArthur and Wilson^{2,15–18}, tests of predictions regarding the dependence of the underlying processes—colonization, speciation and extinction—on island geographical context (area and isolation) are few in number, and are either restricted in temporal,

geographical or taxonomic scope^{5,19,20}, or seek to infer speciation rates in the absence of data on the phylogenetic relationships among species^{2,16}. As a result, there has been no robust and powerful test of MacArthur and Wilson's predictions on a global scale, and the effect of area and isolation on biogeographical processes acting on macro-evolutionary timescales remains largely unexplored.

Here we expand on approaches that leverage the information in time-calibrated molecular phylogenies of insular species^{1,10,21,22} to determine how the processes of colonization, speciation and extinction are influenced by area and isolation. The dynamic stochastic model DAISIE¹⁰ (dynamic assembly of islands through speciation, immigration and extinction) can accurately estimate maximum-likelihood rates of colonization, extinction and speciation rates (CES rates) from branching times (colonization times and any in situ diversification events) and endemism status of species that results from one or multiple independent colonizations of a given island system (for example, all native terrestrial birds on an archipelago)¹⁰. This method can also detect the presence or absence of diversity dependence in rates of colonization and speciation, by estimating a carrying capacity (upper bound to the

¹Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany. ²Naturalis Biodiversity Center, Leiden, The Netherlands. ³Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands. ⁴Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany. ⁵Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. ⁶Museu de História Natural e da Ciência da Universidade do Porto, Porto, Portugal. ⁷Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), InBio, Laboratório Associado, Universidade do Porto, Vairão, Portugal. ⁸FitzPatrick Institute, DST-NRF Centre of Excellence, University of Cape Town, Cape Town, South Africa. ⁹Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, UA, Paris, France. ¹⁰Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK. ¹¹Environmental Futures Research Institute, Griffith University, Brisbane, Queensland, Australia. ¹²Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, Mieres, Spain. ¹³Unité Mixte de Recherche 5174, CNRS-IRD-Paul Sabatier University, Toulouse, France. ✉e-mail: luis.valente@naturalis.nl

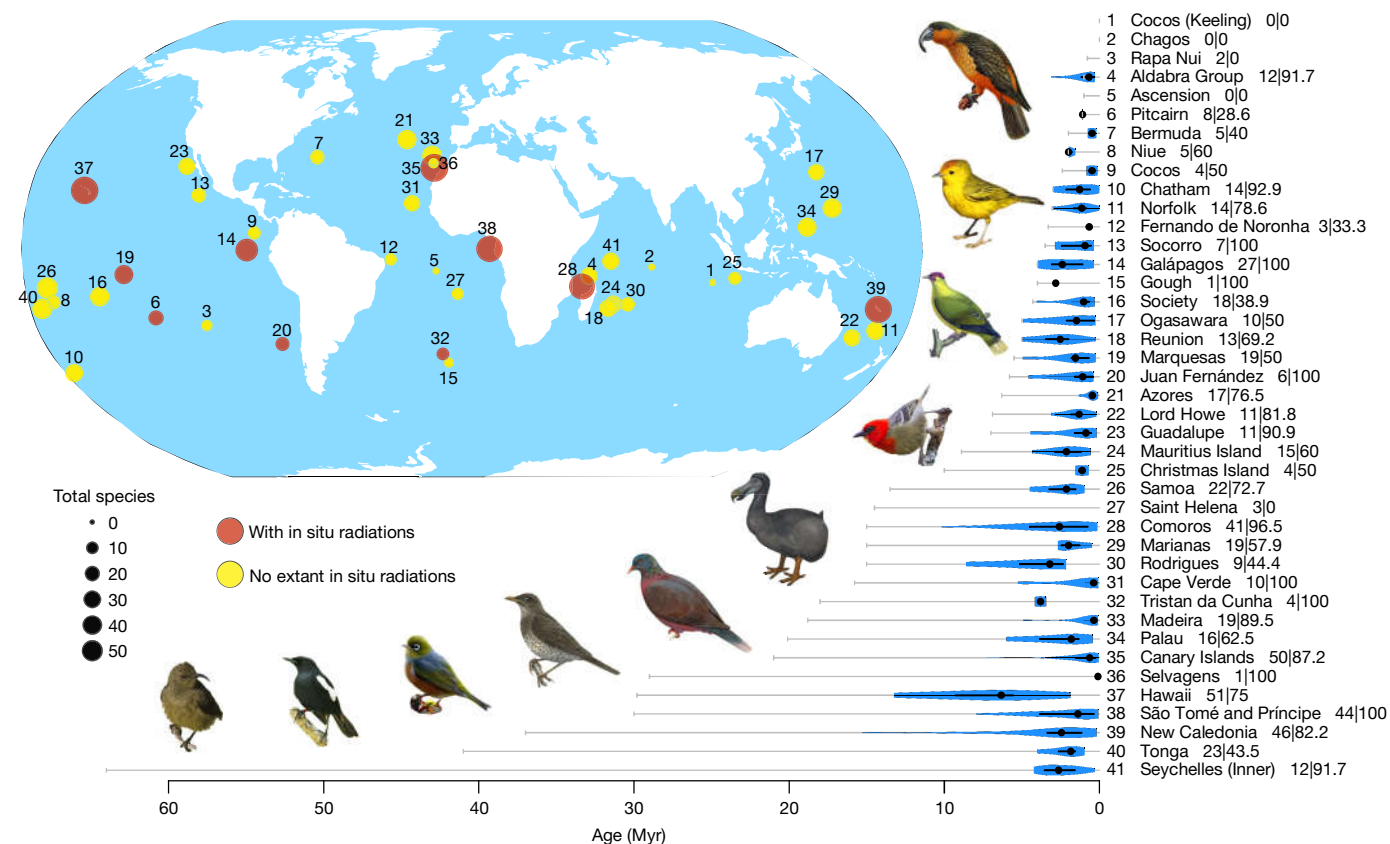


Fig. 1 | Archipelago and island bird colonization time data. Circles show the number of species that belong to our focal group (both extinct and extant) found in each archipelago (at the time of human arrival). Numbers on the map correspond to numbers to the left of the archipelago name. Numbers to the right of the archipelago name indicate the number of species from our focal assemblage on the archipelago | the percentage of species sampled in the phylogenetic trees. Even species not sampled in the trees are accounted for by including them as missing species that could have colonized at any time since the emergence of the archipelago. Colonization times plot: grey horizontal lines indicate archipelago ages (Extended Data Table 1). Violin plots (blue) show the kernel density of the distribution of times of colonization of bird species in each archipelago, obtained from the phylogenetic trees. Thick black lines inside violin plots indicate the interquartile distance; thin black lines indicate

the 95% confidence interval; black dots indicate the median. Archipelagos with no violin plot or dots are cases for which no species of our focal assemblage were present at the time of human arrival, or none were sampled using molecular data. Birds from left to right: Seychelles sunbird, Seychelles magpie robin, silvereye, Príncipe thrush, laurel pigeon, dodo (extinct), Mauritius fody, red-moustached fruit dove (extinct), Galápagos warbler and Norfolk kaka (extinct). Bird images used with permission from: C. Baeta (Príncipe thrush), P. Cascão (Galápagos warbler), M. Hammers (Seychelles sunbird and magpie robin), J. Hume (dodo), D. Shapiro (Mauritius fody) and J. Varela (laurel pigeon). There are no in situ radiations in the Mascarenes (Mauritius, Reunion and Rodrigues) because we treat the islands as separate entities (but see 'Sensitivity to archipelago selection and isolation metrics' in the Methods). Myr, million years.

number of species in an island system). Here we extend DAISIE to estimate the hyperparameters that control the shape of the relationships between CES rates, and the area and isolation of islands worldwide.

The accurate estimation of fundamental island biogeographical relationships requires suitable data from many archipelagos, but divergence-dated phylogenies of complete communities on islands remain scarce. Hence, we produced new dated molecular phylogenies for the terrestrial avifaunas of 41 archipelagos worldwide. Here we refer to both true archipelagos (composed of multiple islands) and isolated insular units that consist of single islands (for example, Saint Helena) as 'archipelago'. For each archipelago, we compiled avian taxon lists (excluding introduced, marine, migratory and aquatic species, as well as birds of prey, rails and nocturnal birds; see Methods) and collected physical data (Fig. 1 and Supplementary Data 1, 2). We use archipelagos as our insular unit, because the high dispersal abilities of birds within archipelagos suggest that, for birds, archipelagos can be considered equivalent to single islands for less dispersive taxa²³, and because archipelagos constitute the most-appropriate spatiotemporal unit for framing analyses of biodiversity patterns at a large scale^{2,24,25}. We extracted colonization and speciation times for each archipelago from

the phylogenetic trees, producing a 'global dataset' for the 41 archipelagos, which includes the complete extant avifauna of each archipelago, plus all species known to have become extinct due to anthropogenic causes. The dataset comprises 596 insular taxa from 491 species. The phylogenies revealed a total of 502 archipelago colonization events and 26 independent in situ 'radiations' (cases in which diversification has occurred within an archipelago), which ranged in size from 2 to 33 species (the Hawaiian honeycreepers being the largest clade). The distribution of colonization times is summarized in Fig. 1 and the full dataset is provided in Supplementary Data 1.

Our extension of the DAISIE framework enables us to estimate hyperparameters that control the relationship between archipelago area and isolation, and archipelago-specific local CES rates, that is, rates of colonization, cladogenesis (within-archipelago speciation that involves in situ lineage splitting), anagenesis (within-archipelago speciation by divergence from the mainland without in situ lineage splitting), natural extinction rates and carrying capacity. We tested the hypothesis that area and distance from the nearest mainland have an effect on the specific CES rates, and, in cases in which a significant effect was identified, estimated its shape and scaling. We developed a set of a priori

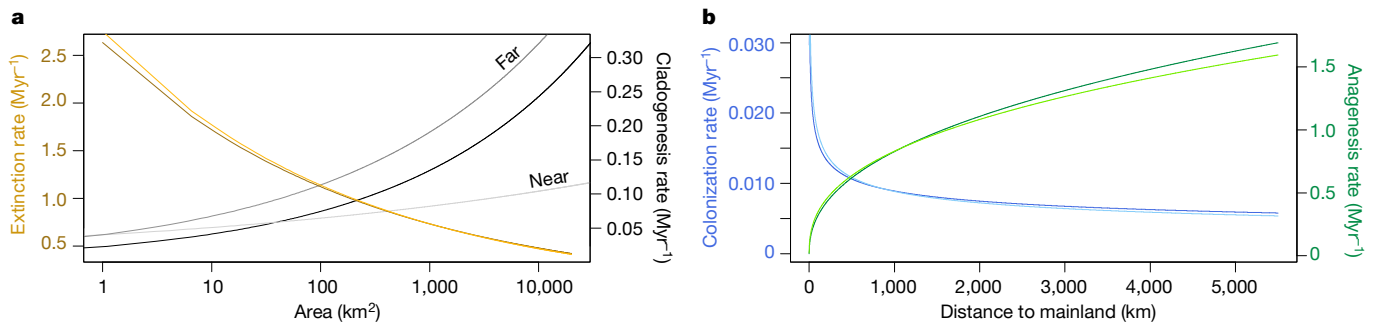


Fig. 2 | Estimated relationships between island area and isolation, and local island biogeography parameters. Isolation was measured as the distance to the nearest mainland. Relationships shown are based on the maximum likelihood global hyperparameters of the best models (equations describing the relationships are provided in Supplementary Table 1). Darker lines, M14 model; lighter lines, M19 model. Under the M14 model, the cladogenesis rate

models (Supplementary Table 1) in which the CES rates are power-law functions of archipelago features. Area has been proposed to have a positive effect on cladogenesis and carrying capacity^{3,5,8} and a negative effect on extinction rates^{8,26}. Archipelago isolation is hypothesized to reduce colonization rates⁷ and increase anagenesis rates²⁷. Models that include or exclude diversity dependence in rates of colonization and cladogenesis¹⁰ (that is, estimating a carrying capacity parameter) were compared. We also considered a set of post hoc models with alternative shapes for the relationships (post hoc power and post hoc sigmoid models; Methods and Supplementary Table 1).

We fitted a set of 28 candidate models to the global dataset using maximum likelihood (Supplementary Table 2). The shape of the relationship of CES rates with area and distance for the two best models is shown in Fig. 2. Under the preferred a priori model (lowest value of the Bayesian information criterion; M14, eight parameters) colonization rates decline with archipelago isolation (exponent of the power law = -0.25 (95% confidence interval = -0.17 – -0.34)) and extinction rate decreases with area (scaling = -0.15 (-0.11 – -0.18)). Rates of cladogenesis increase with area (scaling = 0.26 (0.13 – 0.37)), while anagenesis increases with isolation (scaling = 0.42 (0.24 – 0.61)). The preferred post hoc model (M19, eight parameters) was also the preferred model overall and differs qualitatively from the preferred a priori model M14 only in the cladogenesis function. In the M14 model, cladogenesis is solely a function of area, whereas in the M19 model cladogenesis depends interactively and positively on both area and distance from the nearest mainland, such that the cladogenesis–area relationship is steeper for more isolated archipelagos (Fig. 2 and Extended Data Fig. 1). In addition, we found no evidence for diversity dependence, as the carrying capacity (K) was estimated to be much larger than the number of species on the island and models without a K parameter (no upper bound to diversity), such as M14 and M19, performed better than models that included this parameter (Supplementary Table 2). We also tested whether the inclusion of a combination of true archipelagos and single islands in our dataset could have affected our results, for example if opportunities for allopatric speciation are higher when an area is subdivided into multiple islands²⁸. We repeated analyses in which single island units were excluded and found that the same model (M19) is preferred with similar parameter estimates. We therefore discuss only the results for the main dataset (including both single islands and true archipelagos). Our results are robust to uncertainty in colonization and branching times (see ‘Sensitivity to alternative divergence times and tree topologies’ in the Methods).

A parametric bootstrap analysis of the two preferred models (M14 and M19) demonstrated that the method is able to recover hyperparameters with high precision and little bias (Extended Data Fig. 2). To test the significance of the relationships between area, isolation

depends only on the area. Under the M19 model, the cladogenesis rate increases with both area and distance to the nearest mainland, and thus lines for more (far, 5,000 km) and less (near, 50 km) isolated islands are shown. See Extended Data Fig. 1 for the relationship of cladogenesis with both area and distance under the M19 model.

and CES rates, we conducted a randomization test on the global dataset with reshuffled areas and distances. This test estimated the exponent hyperparameters as zero in most reshuffled cases (that is, no effect of area or isolation was detected; Extended Data Fig. 3), confirming that it is the observed relationships between diversity and archipelago characteristics that generate our parameter estimates.

To assess model fit, we simulated archipelago communities under the best model (M19) and found that—for most archipelagos—the observed diversity metrics (the numbers of species, cladogenetic species and colonizations) were similar to the expected numbers, with some exceptions; for example, diversity was underestimated for Comoros and São Tomé and Príncipe (Fig. 3 and Extended Data Fig. 4). The ability of the model to explain observed values (total species, pseudo- $R^2 = 0.72$; cladogenetic species, pseudo- $R^2 = 0.52$; colonizers, pseudo- $R^2 = 0.60$) was very high considering the model includes only 8 parameters (at least 12 parameters would be needed if each rate depended on area and isolation, and at least 164 parameters if each archipelago was allowed to have its own parameters) and was able to explain multiple diversity metrics. This represents a very large proportion of the explanatory power that would be expected to be obtained for data generated under the preferred model (Extended Data Fig. 5). Simulations under the best model reproduced the classic observed relationships between area, distance and diversity metrics (Fig. 4).

Our approach reveals the empirical shape of fundamental biogeographical relationships that have previously been difficult to estimate. In agreement with recent studies^{2,29}, we found strong evidence for a decline in the rates of colonization with isolation and in the rates of extinction with area, confirming two of the key assumptions of island biogeography theory⁷. The colonization–isolation effect was detected despite the fact that the decline in avian species richness with distance from the nearest mainland in our empirical data was not as pronounced as in other less-mobile taxa^{4,11}, revealing that isolation is a clear determinant of the probability of immigration and the successful establishment of populations even in a highly dispersive group such as birds. The extinction–area relationship has been a fundamental empirical generalization in conservation ecology (for example, for the design of protected areas³⁰); here we were able to characterize the shape of this dependence at the global spatial scale and macro-evolutionary timescale.

We provide insights into the scaling of speciation with area and isolation. In contrast to previous studies on within-island speciation, which have suggested the existence of an area below which cladogenesis does not take place on single islands⁵, we do not find evidence for such an area threshold at the archipelago level and, under our model, speciation is predicted to be non-zero even in small areas. In addition, our post

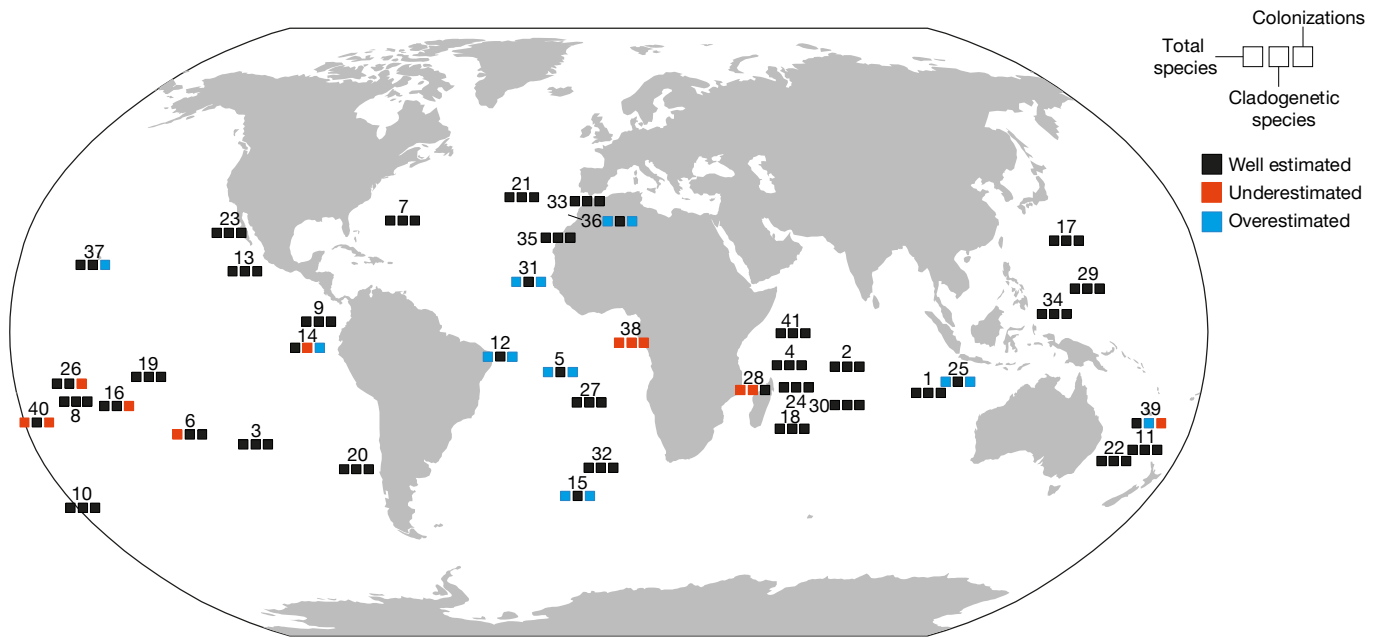


Fig. 3 | Goodness of fit of the preferred model (M19). The map identifies whether the diversity metrics were well estimated (the empirical value matches the 95% confidence interval of simulations), underestimated (the empirical value is higher than the 95% confidence interval) or overestimated (the

empirical value is lower than the 95% confidence interval). Intervals are based on 1,000 simulations of each archipelago (Extended Data Fig. 4). Numbers on the map indicate the archipelagos described in Fig. 1.

hoc finding that rates of cladogenesis increase through an interactive effect of both island size and distance from the nearest mainland (Fig. 2 and Extended Data Fig. 1) provides a mechanism that limits radiations to archipelagos that are both large and remote^{6,27}. Why this interaction exists requires further investigation, but one possibility is that

unsaturated niche space provides greater opportunities for diversification⁶. In addition to the effects of physical features on cladogenesis, we found that rates of anagenesis increase with island isolation. While impressive insular radiations tend to receive the most attention from evolutionary biologists (for example, Darwin's finches or Hawaiian

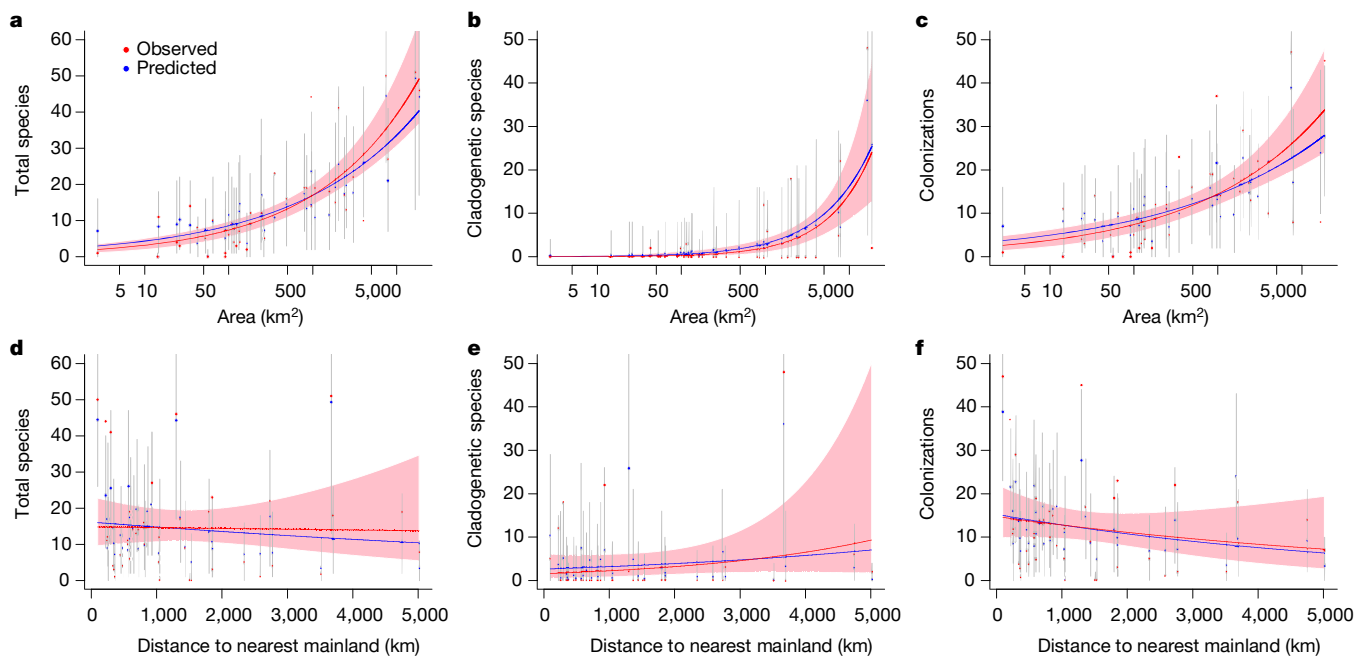


Fig. 4 | Observed and predicted island diversity-area and island diversity-distance relationships. Grey vertical lines show the 95% confidence intervals across 1,000 datasets simulated for each of the 41 archipelagos assuming the M19 model. Blue points indicate the mean values of the simulations; the blue line indicates the fitted line for the simulated data; red points are the observed values in the empirical data; the red line shows the fitted line for the empirical

data; the red shaded area is the 95% confidence interval of the predicted relationship for the empirical data. **a-c**, Relationships between island diversity and area. **a**, Total number of species. **b**, Cladogenetic species. **c**, Number of colonizations. **d-f**, Relationships between island diversity and distance of the island to the mainland. **d**, Total number of species. **e**, Cladogenetic species. **f**, Number of colonizations.

honeycreepers), our phylogenies revealed that the majority of endemic birds in our dataset in fact display an anagenetic pattern (at the time of human arrival, 231 out of 350 endemic species had no extant sister taxa on the archipelago and there were only 26 extant in situ radiations). The positive effect of archipelago isolation on rates of anagenesis that we estimate suggests that this fundamental but overlooked process is impeded by high levels of movement between island and mainland populations.

A variety of global patterns of biodiversity have been described—from small islands and lakes, to biomes and continents—but the processes that underpin these patterns remain to be explored. Our simulations using parameters estimated from data were able to reproduce the classic global patterns of island biogeography across 41 archipelagos (Fig. 4). This advances our understanding of macro-scale biology, by providing missing links between local processes, environment and global patterns. More than half a century after the seminal work of MacArthur and Wilson⁷, we now have the data and tools to go beyond statistical descriptions of diversity patterns, enabling us to quantify community-level processes that have long been unclear.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2022-5>.

- Ricklefs, R. E. & Bermingham, E. Nonequilibrium diversity dynamics of the Lesser Antillean avifauna. *Science* **294**, 1522–1524 (2001).
- Triantis, K. A., Economo, E. P., Guilhaumon, F. & Ricklefs, R. E. Diversity regulation at macro-scales: species richness on oceanic archipelagos. *Glob. Ecol. Biogeogr.* **24**, 594–605 (2015).
- Whittaker, R. J., Triantis, K. A. & Ladle, R. J. A general dynamic theory of oceanic island biogeography. *J. Biogeogr.* **35**, 977–994 (2008).
- Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. Global diversity of island floras from a macroecological perspective. *Ecol. Lett.* **11**, 116–127 (2008).
- Losos, J. B. & Schluter, D. Analysis of an evolutionary species–area relationship. *Nature* **408**, 847–850 (2000).
- Gillespie, R. G. & Baldwin, B. G. In *The Theory of Island Biogeography Revisited* (eds Losos, J. & Ricklefs, R. E.) 358–387 (Princeton Univ. Press, 2010).
- MacArthur, R. H. & Wilson, E. O. An equilibrium theory of insular zoogeography. *Evolution* **17**, 373–387 (1963).
- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
- Warren, B. H. et al. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur–Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
- Valente, L. M., Phillimore, A. B. & Etienne, R. S. Equilibrium and non-equilibrium dynamics simultaneously operate in the Galápagos islands. *Ecol. Lett.* **18**, 844–852 (2015).
- Lomolino, M. V. Species–area and species–distance relationships of terrestrial mammals in the Thousand Island Region. *Oecologia* **54**, 72–75 (1982).
- Diamond, J. M. Biogeographic kinetics: estimation of relaxation times for avifaunas of southwest Pacific islands. *Proc. Natl Acad. Sci. USA* **69**, 3199–3203 (1972).
- Whittaker, R. J. & Fernandez-Palacios, J. M. *Island Biogeography: Ecology, Evolution, and Conservation* (Oxford Univ. Press, 2007).
- Matthews, T. J., Rigal, F., Triantis, K. A. & Whittaker, R. J. A global model of island species–area relationships. *Proc. Natl Acad. Sci. USA* **116**, 12337–12342 (2019).
- Weigelt, P., Steinbauer, M. J., Cabral, J. S. & Kreft, H. Late Quaternary climate change shapes island biodiversity. *Nature* **532**, 99–102 (2016).
- Lim, J. Y. & Marshall, C. R. The true tempo of evolutionary radiation and decline revealed on the Hawaiian archipelago. *Nature* **543**, 710–713 (2017).
- Cabral, J. S., Weigelt, P., Kissling, W. D. & Kreft, H. Biogeographic, climatic and spatial drivers differentially affect α -, β - and γ -diversities on oceanic archipelagos. *Proc. R. Soc. B* **281**, 20133246 (2014).
- Matthews, T. J., Guilhaumon, F., Triantis, K. A., Borregaard, M. K. & Whittaker, R. J. On the form of species–area relationships in habitat zoogeography and true islands. *Glob. Ecol. Biogeogr.* **25**, 847–858 (2016).
- Simberloff, D. S. & Wilson, E. O. Experimental zoogeography of islands: the colonization of empty islands. *Ecology* **50**, 278–296 (1969).
- Russell, G. J., Diamond, J. M., Reed, T. M. & Pimm, S. L. Breeding birds on small islands: island biogeography or optimal foraging? *J. Anim. Ecol.* **75**, 324–339 (2006).
- Rabosky, D. L. & Glor, R. E. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proc. Natl Acad. Sci. USA* **107**, 22178–22183 (2010).
- Emerson, B. C. & Gillespie, R. G. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.* **23**, 619–630 (2008).
- Kisel, Y. & Barraclough, T. G. Speciation has a spatial scale that depends on levels of gene flow. *Am. Nat.* **175**, 316–334 (2010).
- Triantis, K., Whittaker, R. J., Fernández-Palacios, J. M. & Geist, D. J. Oceanic archipelagos: a perspective on the geodynamics and biogeography of the World’s smallest biotic provinces. *Front. Biogeogr.* **8**, e29605 (2016).
- Santos, A. M. C. et al. Are species–area relationships from entire archipelagos congruent with those of their constituent islands? *Glob. Ecol. Biogeogr.* **19**, 527–540 (2010).
- Ricklefs, R. E. & Lovette, I. J. The roles of island area per se and habitat diversity in the species–area relationships of four Lesser Antillean faunal groups. *J. Anim. Ecol.* **68**, 1142–1160 (1999).
- Rosindell, J. & Phillimore, A. B. A unified model of island biogeography sheds light on the zone of radiation. *Ecol. Lett.* **14**, 552–560 (2011).
- Losos, J. B. & Ricklefs, R. E. Adaptation and diversification on islands. *Nature* **457**, 830–836 (2009).
- Keil, P. et al. Spatial scaling of extinction rates: theory and data reveal nonlinearity and a major upscaling and downscaling challenge. *Glob. Ecol. Biogeogr.* **27**, 2–13 (2016).
- Wilcox, B. A. & Murphy, D. D. Conservation strategy: the effects of fragmentation on extinction. *Am. Nat.* **125**, 879–887 (1985).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Archipelago selection

We focus on oceanic islands, that is, volcanic islands that have never been connected to any other landmass in the past. We also include the Granitic Inner Seychelles, even though these islands have a continental origin, because they have been separated from other landmasses for a very long period of time (64 million years)³¹ and can be considered quasi-oceanic, as all extant avian species originated in much more recent times. The 41 archipelagos chosen are located in the Atlantic, Indian and Pacific Oceans, with latitudes between 45° north and south. Islands within these archipelagos are separated by a maximum of 150 km. The sole exceptions are the Azores and Hawaii, two very isolated systems where the distances between some islands exceed this value. The shape files used to plot the maps of Figs. 1, 3 were obtained from a previous study³².

Physical and geological data

Full archipelago data are provided in Supplementary Data 2 and Extended Data Table 1. We obtained data on the total contemporary landmass area for each archipelago. For our isolation metric, we computed the minimum round Earth distance to the nearest mainland (D_m) in km using Google Earth. We considered 'nearest mainland' to be the nearest probable source of colonists (but see 'Sensitivity to archipelago selection and isolation metrics' for different isolation metrics). This is the nearest continent except for island groups that were closer to Madagascar, New Guinea or New Zealand than to the continent, in which case we assigned these large continent-like islands as the mainland. This is supported by our phylogenetic data—for example, many Indian Ocean island taxa have closest relatives on Madagascar rather than mainland Africa.

Island palaeo-areas and past archipelago configurations have been shown to be better predictors of endemic insular diversity than contemporary area^{15,33}. By contrast, island total native and non-endemic richness is better predicted by present island characteristics^{15,33}. As insufficient data on island ontogeny was available (that is, describing the empirical area trajectories from island birth to present), we analysed contemporary area and isolation as these are currently the most appropriate units for our dataset.

We conducted an extensive survey of the literature and consulted geologists to obtain the geological ages for each archipelago (Extended Data Table 1), treating the age of the oldest currently emerged island as an upper bound for colonization. Islands may have been submerged and have emerged multiple times and we consider the age of the last known emergence. For the Aldabra Group we used an age older than the published estimate. The current estimated age of re-emergence of Aldabra is 0.125 million years³⁴, but 9 out of 12 Aldabra colonization events in our dataset are older, suggesting that the archipelago was not fully submerged before this and may have been available for colonization for a longer period. Therefore, for Aldabra we used an older upper bound of 1 million years for colonization, although we acknowledge that the mitochondrial markers used for dating may not provide sufficient resolution at the shallow temporal scale of the published age. For Hawaii, the colonization times that we obtained for more than half of the colonization events were older than the age of the current high islands that is often used as a maximum age for colonization (around 5 million years). Therefore, instead of this age, we used the much older estimate of 29.8 million years of the Kure Atoll³⁵ to account for currently submerged or very low-lying Hawaiian Islands that could have received colonists in the past. For Bermuda and Marianas, we could not find age estimates in the literature, and we therefore consulted geologists to obtain these (P. Hearty, R. Stern and M. Reagan, personal communication; Extended Data Table 1).

Island avifaunas

Our sampling focused on native resident terrestrial birds and we considered only birds that colonize by chance events (for example, hurricanes

or rafts). We thus excluded marine and migratory species, because they are capable of actively colonizing an island at a much higher rate. We focused on songbird-like and pigeon-like birds, which constitute the majority of terrestrial (land-dwelling) birds on islands. Following a precedent set by previous work^{10,27,36}, we included only species from the same trophic level (in the spirit of MacArthur and Wilson's model): we excluded aquatic birds, birds of prey, rails (many are flightless or semi-aquatic) and nightjars (nocturnal). We also excluded introduced and vagrant species. Including species such as rails and owls (which are components of many island avifaunas) would have led to a higher estimate of the product of colonization rate and mainland pool size due to a larger mainland pool, and potentially to higher estimated rates of anagenesis (many owl or rail species are island endemics with no close relatives on the islands).

For the focal avian groups, we compiled complete taxon lists for each of the 41 archipelagos based on recent checklists from Avibase (<http://avibase.bsc-eoc.org>), which we cross-checked with the online version of the *Handbook of the Birds of the World* (HBW)³⁷. We followed the HBW's nomenclature and species assignments, except for 12 cases in which our phylogenetic data disagree with HBW's scheme (noted in the column 'Taxonomy' of Supplementary Data 1). For example, in 11 cases phylogenetic trees support raising endemic island subspecies to species status (we sampled multiple samples per island taxon and outgroup, and the island individuals form a reciprocally monophyletic well-supported clade), and for these taxa we decided it was more appropriate to use a phylogenetic species concept so as not to underestimate endemism and rates of speciation (Supplementary Data 1). We re-ran DAISIE analyses using HBW's classification and found that the maximum-likelihood parameters are very similar and thus we report only the results using the scheme based on the phylogenies produced for this study.

For each bird species found on each archipelago, we aimed to sample sequence data for individuals on the archipelago and the closest relatives outside the archipelago (outgroup taxa). Our sampling success per archipelago is shown in Fig. 1 and Extended Data Table 1.

Extinct species

We do not count extinctions with anthropogenic causes as influencing the natural background rate of extinction. Therefore, we explicitly include species for which there is strong evidence that they have been extirpated by humans. We treat taxa extirpated on an archipelago by humans as though they had survived in that archipelago until the present following our previously published approach³⁸.

We identified anthropogenic extinctions based on published data^{39–46} and personal comments (J. A. Alcover and J. C. Rando on unpublished Macaronesian taxa; F. Sayol and S. Faurby). We include the species present on the islands that belong to our archipelago definition as described in Supplementary Data 2. We excluded largely hypothetical accounts or pre-Holocene fossils that greatly predate human arrival. Our dataset accounts for 153 taxa that were present on first human contact and have gone extinct since, probably because of human activities including the introduction of invasive species by humans. To our knowledge, 71 of these taxa have previously been sequenced using ancient DNA or belong to clades present in our trees, and we were thus able to include them in the phylogenetic analyses as regular data ($n = 54$), or as missing species by adding them as unsampled species to a designated clade ($n = 17$). For the remaining 82 extinct taxa, sequences were not available and we were unable to obtain samples and to allocate them to clades. We assume that these taxa represent extinct independent colonizations and we included them in the analyses using the 'Endemic_MaxAge' and 'Non_endemic_MaxAge' options in DAISIE, which assume that they have colonized at any given time since the birth of the archipelago (but before any *in situ* cladogenesis event). As an example, our dataset includes the 27 species of Hawaiian birds belonging to our focal group that are known to have gone extinct since human

Article

colonization. Eight of these species were included using DNA data, 17 were added as missing species to their clades (14 honeycreepers and 3 *Myadestes*) and two were added using the Endemic_MaxAge option in DAISIE (*Corvus impluviatus* and *Corvus viriosus*).

Sequence data from GenBank

We conducted an extensive search of GenBank for available DNA sequences from the 596 island bird taxa that fitted our sampling criteria and from multiple outgroup taxa, using Geneious v.11⁴⁷. The molecular markers chosen varied from species to species, depending on which marker was typically sequenced for the taxon in question, the most commonly sequenced marker was cytochrome *b*. In total, we downloaded 3,155 sequences from GenBank. For some taxa, sequences from both archipelago and close relatives from outside the archipelago were already available from detailed phylogenetic or phylogeographical analyses. In some cases, a target species had been sampled, but only from populations outside the archipelago. In other cases, the species on the archipelago had been sampled, but the sampling of the relatives outside of the archipelago was lacking or only available for distant regions, which meant a suitable outgroup was not available in GenBank. Finally, for some species there were no previous published sequences available in GenBank. GenBank accession numbers and geographical origin for the downloaded sequences are provided in the DNA matrices (<https://doi.org/10.17632/vf95364vx6.1>) and maximum clade credibility trees (<https://doi.org/10.17632/p6hm5w8s3b.2>) uploaded to Mendeley Data.

Sequence data of new samples

Sequences available in GenBank covered only 54% (269 out of 502) of the total independent colonization events. We improved the sampling by obtaining new sequences for many island taxa ($n = 174$ taxa) and from their close relatives from continental regions ($n = 78$). We obtained new samples from three sources: field trips, research collections and colleagues who contributed field samples. New samples were obtained during field trips conducted by M.M. (Gulf of Guinea and African continent); B.H.W. and C.T. (Comoros and Mayotte, Mauritius Island, Rodrigues, Seychelles); S.M.C. (New Caledonia); J.C.I. (Macaronesia, Europe and Africa) and L.V. (New Caledonia), between 1999 and 2017. Samples of individuals were captured using mist-nets or spring traps baited with larvae. Blood samples were taken by brachial venipuncture, diluted in ethanol or Queen's lysis buffer in a micro-centrifuge tube. Birds were released at the point of capture. Aldabra Group samples were obtained from research collections of the Seychelles Islands Foundation. Museum samples from several Galápagos and Comoros specimens were obtained on loan from, respectively, the California Academy of Sciences and the Natural History Museum London. Additional samples from various localities (Aldabra Islands, Iberian Peninsula, Madagascar and Senegal) were provided by collaborators, as indicated in Supplementary Table 3. Sample information and GenBank accession numbers for all new specimens are provided in Supplementary Table 3.

DNA was extracted from blood, feathers and museum toe-pad samples using QIAGEN DNeasy Blood and Tissue kits (Qiagen). For museum samples, we used a dedicated ancient DNA laboratory facility at the University of Potsdam to avoid contamination. The cytochrome *b* region (1,100 base pairs) was amplified using the primers shown in Extended Data Table 2. DNA from historical museum samples was degraded and cytochrome *b* could not be amplified as a single fragment. We thus designed internal primers to sequence different overlapping fragments in a stepwise manner (Extended Data Table 2).

PCRs were set up in 25- μ l total volumes including 5 μ l of buffer Boline MyTaq, 1 μ l (10 mM) of each primer and 0.12 μ l MyTaq polymerase. PCRs were performed with the following thermocycler conditions: initial denaturation at 95 °C for 1 min followed by 35 cycles of denaturation at 95 °C for 20 s, with an annealing temperature of 48 °C for 20 s, and extension at 72 °C for 15 s and a final extension at 72 °C for 10 min.

Amplified products were purified using exonuclease I and Antarctic phosphatase, and sequenced at the University of Potsdam (Unit of Evolutionary Biology/Systematic Zoology) on an ABI PRISM 3130xl sequencer (Applied Biosystems) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). We used Geneious v.11 to edit chromatograms and align sequences.

Phylogenetic analyses

To estimate times of colonization and speciation for each archipelago, we produced new divergence dated phylogenies or compiled published dated trees, to yield a total of 91 independent phylogenies (maximum clade credibility trees and posterior distribution deposited in Mendeley, <https://doi.org/10.17632/p6hm5w8s3b.2>) for all new trees produced for this study; the 11 previously published trees are available upon request). Information on all alignments and trees, including molecular markers, data sources, calibration methods and substitution model are provided in Extended Data Tables 3, 4 and Supplementary Table 4. The majority of alignments and phylogenies focus on a single genus, although some include multiple closely related genera or higher order clades (family, order) depending on the diversity and level of sampling of the relevant group (taxonomic scope is described in Extended Data Tables 3, 4). Most alignments include taxa from a variety of archipelagos. Alignments were based on a variety of markers, according to which marker had most often been sequenced for a given group.

For the new dating analyses conducted for this study, we created 80 separate alignments for different groups using a combination of sequences from GenBank ($n = 3,155$) and new sequences ($n = 252$) produced for this study. In some cases, we obtained DNA alignments directly from authors of previous studies and these are credited in Extended Data Table 3. Phylogenetic divergence dating analyses were performed in BEAST 2⁴⁸. For each alignment, we performed substitution model selection in jModeltest⁴⁹ using the Bayesian information criterion (BIC). We used rates of molecular evolution for avian mitochondrial sequences, which have been shown to evolve in a clock-like manner at an average rate of around 2% per million years⁵⁰. Molecular rate calibrations can be problematic for ancient clades, due to high levels of heterotachy in birds⁵¹. In addition, mitochondrial DNA saturates after about 10–20 million years, and genetic distances of more than 20% may provide limited information regarding dating⁵². Therefore, we only used molecular rate dating to extract node ages for branching events at the tips of the trees, at the species or population level (oldest colonization time in our dataset is 15.3 million years, but most are much younger). Rates of evolution were obtained from the literature and varied between different markers and taxonomic group (Supplementary Table 4). We applied the avian mitochondrial rates estimated using cytochrome *b* from a previous study⁵⁰ (but see 'Sensitivity to alternative divergence times and tree topologies' for different rates).

We applied a Bayesian uncorrelated log-normal relaxed clock model. For each analysis, we ran two independent chains of between 10 and 40 million generations, with a birth–death tree prior. We assessed convergence of chains and appropriate burn-ins with Tracer, combined runs using LogCombiner and produced maximum clade credibility trees with mean node heights in Tree Annotator. We produced a total of 80 maximum clade credibility trees.

For 11 groups (Extended Data Table 4), well-sampled and rigorously dated phylogenies were already available from recent publications, all of which conducted Bayesian divergence dating using a variety of calibration methods, including fossils and molecular rates. We obtained maximum clade credibility trees from these studies from online repositories or directly from the authors (Extended Data Table 4).

Colonization and branching times

The nodes selected in the dated trees for estimates of colonization and branching times are given for each taxon in Supplementary Data 1. Our node selection approach was as follows. For cases in which samples

representing species or populations from archipelagos formed a monophyletic clade consisting exclusively of archipelago individuals, we used the stem age of this clade as colonization time. For cases in which only one individual of the archipelago was sampled, we used the length of the tip leading to that individual, which is equivalent to the stem age. For cases in which the archipelago individuals were embedded in a clade containing mainland individuals of the same species—that is, paraphyly or polyphyly—we assumed (based on morphological characteristics) that this is due to incomplete lineage sorting of the insular and mainland lineages, and we therefore used the most recent common ancestor node of the archipelago individuals, or the crown node when the most recent common ancestor node coincides with the crown. For these later cases, using the stem would most likely have been an overestimation of the colonization time, as we assume that colonization happens from the mainland to the archipelago. For such cases, we applied the ages using the ‘MaxAge’ option in DAISIE, which integrates over the possible colonization times between the present and the upper bound. A robustness test of our results to node choice is given in ‘Sensitivity to alternative branching times and tree topologies’.

For a total of 19 endemic taxa we could not obtain sequences, but we could allocate them to a specific island clade (for example, Hawaiian honeycreepers and solitaires). These were added as missing species to that clade. For 96 non-endemic taxa we could not obtain sequences of individuals from the archipelago, but we could obtain sequences from the same species from different regions. For these cases, we used the crown or the stem age of the species as an upper bound for the age of the colonization event, using the ‘Non_endemic_MaxAge’ option in DAISIE. Finally, for 124 taxa (20.8%) no sequences of individuals from the archipelago were available in GenBank and we were not able to obtain samples for sequencing from the species or from close relatives. We assumed these cases constituted independent colonizations that could have taken place any time since the origin of the archipelago and the present, and applied the ‘Non_endemic_MaxAge’ and ‘Endemic_MaxAge’ options in DAISIE with a maximum age equal to the archipelago age. DAISIE makes use of the information described above; further information has been described previously⁵³.

Global dataset characteristics

Data points from taxa of the same archipelago were assembled into 41 archipelago-specific datasets. These 41 datasets were in turn assembled into a single dataset (D1), which was analysed with DAISIE (D1 DAISIE R object, available in Mendeley Data <https://doi.org/10.17632/sy58zbv3s2.2>). This dataset (Supplementary Data 1) has a total of 596 taxa (independent colonization events plus species within radiations), covering 491 species from 203 different genera and 8 orders. All taxa were included in the analyses: not only those which we sampled in phylogenies, but also those for which sequences or phylogenies could not be obtained and which were included following the approaches described in ‘Colonization and branching time’. A summary of diversity and sampling per archipelago is provided in Extended Data Table 1.

Sampling completeness

In total, we produced new sequences from 252 new individuals, comprising 90 different species from 45 different genera, covering an additional 110 colonization events that had not been sampled (that is, populations from islands for which the species had not been sampled before). For at least 12 of these 90 species, we found no previous sequences in GenBank, including island endemics from Comoros, Galápagos, Rodrigues and São Tomé (Supplementary Table 5). The new sequences from 252 individuals increase the molecular sampling for extant colonization events from 60% (223 out of 373) to 89% (332 out of 373). If we include historically extinct colonizations, we increased the molecular sampling from the existing 54% (269 out of 502) of colonization events to 75% (379 out of 502). We also substantially increased molecular sampling of continental relatives, adding 78 new individuals from the continent

or islands surrounding our archipelagos, covering 43 different species. The percentage of taxa sampled in phylogenies varied widely between archipelagos (Fig. 1 and Extended Data Table 1). For 8 archipelagos (Bermuda, Fernando de Noronha, Pitcairn, Rapa Nui, Rodrigues, Saint Helena, Society Islands and Tonga) less than 50% of the species were sampled in phylogenies, and thus the majority of the species for these island groups were added with maximum ages and endemism status. For 13 archipelagos, which accounted for more than a third of the total species, over 90% of the species were sampled in phylogenies.

DAISIE

We used the method DAISIE¹⁰ to estimate rates of species accumulation (colonization, speciation and extinction) on the archipelagos. The model assumes that after the origin of an island, species can colonize from a mainland pool. Once a species has colonized, it may remain similar to its mainland ancestor (non-endemic species), become endemic through anagenetic speciation (new endemic species is formed without lineage splitting on the island), split into new species via cladogenetic speciation and/or go extinct. A carrying capacity (that is, the maximum number of species each colonist lineage can attain) is implemented, such that rates of cladogenesis and colonization decline with increasing number of species in the colonizing clade.

The only effect of anagenesis in DAISIE is that the colonizing species becomes endemic, because further anagenesis events on the endemic species do not leave a signature in the data. However, the rate of anagenesis is not systematically underestimated. Suppose the rate was higher; it would then follow that colonizing species would also become endemic faster, and we would see more endemic species. Thus, the number of endemic species determines the rate of anagenesis, and DAISIE estimates the true rate of anagenesis without systematic bias. Further anagenesis events do not have an effect on the state variables, and hence do not enter the equations anymore.

In its parameterization of extinction, DAISIE accounts for the fact that there may have been several lineages that were present on the insular system in the past but that went completely extinct due to natural causes, leaving no extant descendants. Simulations have shown that the rate of natural extinction is usually well estimated in DAISIE (see ‘Measuring precision and accuracy’ and a previously published study⁵³). Studies on phylogenies of single clades suggest that phylogenetic data on only extant species provide less information on extinction than on speciation (or rather diversification rates⁵⁴). However, there is information content in such data⁵⁵, especially when diversification dynamics are dependent on diversity⁵⁶. Moreover, here we use colonization times in addition to phylogenetic branching times to estimate extinction rates, and we are estimating hyperparameters of the theoretically and empirically suggested relationship of extinction with area. Finally, we use data from many independent colonizations, which increases the power of our statistical method considerably and decreases the bias, as maximum likelihood is known to asymptotically provide unbiased estimates.

Estimating global hyperparameters

Our aim is to examine the dependencies of the parameters that govern species assembly (colonization, extinction, cladogenesis, anagenesis (CES rates) and carrying capacity) on the features of archipelagos (area and isolation). We developed a method to estimate global hyperparameters that control the relationship between two key archipelago features (area and isolation) and archipelago-specific (local) CES rates. One can estimate directly from the global dataset the shape of the relationship between isolation and colonization rate that maximizes the likelihood for the entire dataset.

Our method finds the hyperparameters that maximize the likelihood of the entire dataset, that is, the sum of the log-likelihoods for each archipelago. We tested the hypothesis that area and distance from the nearest mainland have an effect on CES rates (cladogenesis, anagenesis,

Article

extinction and colonization). If an effect was identified, we also estimated the scaling of the effect. We developed a set of a priori models in which the CES rates are affected by archipelago features as is often assumed in the island biogeography literature (Supplementary Table 1). For the a priori models, we considered that CES rates are determined by a power function of area or distance. In the power function, $\text{par} = \text{par}_0 I^h$ where par is the CES rate (for example, local rate of colonization), par_0 is the initial value of the biogeographical rate (for example, global initial rate of colonization), I is the physical variable (area or distance) and h is the strength of the relationship. The exponent h can be negative or positive depending on the nature of the relationship. par_0 and h are the hyperparameters. If the exponent h is estimated as zero, there is no relationship between I and the parameter. By including or excluding h from the different relationships, we can compare different models with the effects switched on or off (Supplementary Table 1; for example, in model M1 all relationships are estimated, whereas in model M2 the exponent of the relationship between anagenesis and distance is fixed to zero and thus anagenesis does not vary with distance).

In addition to the a priori models, we considered a set of post hoc models with alternative shapes of relationships. We fitted two types of post hoc models: power models and sigmoid models (Supplementary Table 1). In the post hoc power models, we modelled all parameters as in the a priori models, except for cladogenesis: we allowed cladogenesis to be dependent on both area and distance. The reason for this is that we found that the predicted number of cladogenetic species under the a priori models were not as high as observed, so we examined whether including a positive effect of distance would improve the fit. We described the relationship between area, distance and cladogenesis using different functions—one model in which there is an additive effect of area and distance (M15); and three models (M16, M17 and M18) in which the effect of area and distance is interactive. In addition, we fitted a model identical to M16 but with one parameter less (M19). The reason for this was that this parameter (γ) was estimated to be zero in M16.

In the post hoc sigmoid models, we allowed the relationship between distance and a given parameter to follow a sigmoid rather than a power function. The rationale for this was that we wanted to investigate whether, for birds, the effect of distance on a parameter only starts to operate after a certain distance from the mainland, as below certain geographical distances archipelagos are within easy reach for many bird species by flight, so that at these distances the island behaves almost as part of the mainland from a bird's perspective. We fitted nine different sigmoid models (Supplementary Table 1), allowing cladogenesis, anagenesis and colonization to vary with distance following a sigmoid function. The sigmoid function that we used has an additional parameter in comparison to power functions.

In total, we fitted 28 candidate models (14 a priori, 14 post hoc) to the global dataset using maximum likelihood. We fitted each model using 20 initial sets of random starting parameters to reduce the risk of being trapped in local likelihood suboptima. We used the age of each archipelago (Extended Data Table 1) as the maximum age for colonization. We assumed a global mainland species pool M of 1,000 species. The product of M and the intrinsic rate of colonization (γ_0) is constant as long as M is large enough (larger than the number of island species), and thus the chosen value of M does not affect the results.

To decide which information criterion to use to select between different models, we compared the performance of the BIC and the Akaike information criterion (AIC). We simulated 1,000 datasets each with models M9 and M19 and then fitted the M9, M14, M17 and M19 models to each of these datasets using two initial sets of starting parameters for each optimization. We found that for datasets simulated using M9 an incorrect model was preferred using AIC in 10.4% of cases, but only in 0.11% of cases when using BIC. For datasets simulated using M19 an incorrect model was preferred 12.8% of cases using AIC and 11.1% of cases using BIC. We thus compared models using BIC, as this model has lower error rates.

An alternative approach to estimating hyperparameters would be to calculate CES rates and their uncertainty independently for each archipelago and to then conduct a meta-analysis of the resulting data, including archipelago area and isolation as predictors. However, errors in parameter estimates will vary, particularly because some archipelagos have small sample sizes (only a few extant colonization events, or none at all; for example, Chagos) and are thus much less informative about underlying process⁵³. Thus, maximizing the likelihood of all datasets together by estimating the hyperparameters (which is precisely our aim) is preferable. For completeness, we present CES rates estimated independently for each archipelago in Supplementary Table 6, excluding archipelagos with fewer than 6 species and for which we sampled less than 60% of the species in the phylogenies. However, as argued above we do not advocate using these parameter estimates for further analyses because the number of taxa for some of these archipelagos is still low and by excluding archipelagos with fewer than six taxa, we cannot capture the lower part of the relationship between area or isolation and CES rates.

All DAISIE analyses were run using parallel computation on the high-performance computer clusters of the University of Groningen (Peregrine cluster) and the Museum für Naturkunde Berlin. The new version of the R package DAISIE is available on GitHub.

Randomization analysis

We conducted a randomization analysis to evaluate whether there is significant signal of a relationship between area and distance and local CES rates in our global dataset. We produced 1,000 datasets with the same phylogenetic data and archipelago ages as the global dataset, but randomly reshuffled archipelago area and D_m in each dataset. We then fitted the best post hoc model to each of these 1,000 randomized datasets. If the maximum-likelihood estimates of exponent hyperparameters (that is, the strength of the relationship) in the randomized datasets were non-zero, this would indicate that the method is finding evidence for a relationship even if there is none. If, on the other hand, non-zero hyperparameters are estimated in the real data but not in the randomized datasets, this would mean that there is information in the data regarding the putative relationships.

The randomization analysis showed that in global datasets with reshuffled areas and distances the exponent hyperparameters are estimated as zero in most cases, whereas in the empirical global dataset they are not (Extended Data Fig. 3).

A posteriori simulations

We simulated 1,000 phylogenetic global datasets (41 archipelagos each) with the maximum-likelihood hyperparameters of the best a priori (M14) and post hoc models (M19). We first calculated the local CES rates for each archipelago based on their area and isolation and the hyperparameters for the model, and then used these CES rates as the parameters for the simulations using the DAISIE R package. The simulated data were used to measure bias and accuracy of the method, goodness of fit and the ability of our method to recover observed island biogeographical diversity patterns (see 'Measuring precision and accuracy of method' and 'Measuring goodness of fit' sections).

Measuring precision and accuracy of method

DAISIE estimates the CES rates with high precision and little bias^{10,53}. We conducted parametric bootstrap analyses to assess whether the ability to estimate hyperparameters from global datasets is also good (Extended Data Fig. 2) and to obtain confidence intervals on parameter estimates (Extended Data Table 5). We used DAISIE to estimate hyperparameters from the M14 and M19 simulated datasets (1,000 replicates each). We measured precision and accuracy by comparing the distribution of parameters estimated from the 1,000 simulated dataset with the real parameters used to simulate the same datasets. To check whether maximum-likelihood optimizations of the simulated

global datasets converge to the same point in parameter space, we first performed a test on a subset of the simulated data. We ran optimizations with 10 random sets of initial starting values for each of 10 simulated datasets. All optimizations converged to the same likelihood and a very similar hyperparameter set; therefore, we are confident that we found the global optimum for each simulated global dataset, even for models with many parameters.

Measuring goodness of fit

We measured how well the preferred models fitted the data using different approaches. First, we examined whether our models successfully reproduced the diversity patterns found on individual archipelagos. We calculated the total number of species, cladogenetic species and independent colonizations in each archipelago for each of the 1,000 simulated datasets. We then plotted these metrics versus the observed values in the empirical data (Fig. 3 and Extended Data Fig. 4). Our preferred models have a slight tendency to overpredict species richness when there are a few species and underpredict it when there are many. We do not have a clear explanation for this. This slight deviation does not seem to be due to an additional dependence on area or distance, so an explanation should be sought in other factors that we did not model. We note that the fact that all three plots show this tendency rather than only one is to be expected because the three metrics of species richness are not entirely independent, with total species richness being the sum of the other two.

Second, we examined whether the models successfully predict the empirical relationships between area, distance and diversity metrics (total species, cladogenetic species, and number of independent colonizations). We fitted generalized linear models for each diversity metric, with quasi-Poisson family errors and log area (or distance) as predictors. We then repeated this across 1,000 independent sets of simulated data for the 41 archipelagos and compared the mean of slopes and intercepts for archipelago area and archipelago isolation to the equivalent estimates for the empirical data (Fig. 4).

Third, we estimated the pseudo- R^2 of the best model (M19) as a measure of the explanatory power of the model. We simulated two independent sets of 10,000 global datasets under M19 model (set 1 and set 2). We calculated the mean total number of species, number of cladogenetic species and colonizations for each archipelago across all datasets from set 1. For each diversity metric, we calculated a pseudo- R^2 (pseudo- R^2 observed) for which the total sum of squares was obtained from the empirical data and the residual sum of squares was calculated as the difference between empirical values and expected values (that is, the simulation means). As the model is inherently stochastic, even if the model is an accurate and complete reflection of the underlying processes then the pseudo- R^2 would tend to be <1. To estimate the distribution of pseudo- R^2 expected under the model, we treated the set-2 simulations as data and estimated the pseudo- R^2 for each (pseudo- R^2 simulated). We then calculated the ratio of the pseudo- R^2 -observed values over the 10,000 pseudo- R^2 -simulated values. A ratio approaching 1 would indicate that the model is explaining the observed data as well as the average dataset simulated under this process (Extended Data Fig. 5).

Sensitivity to alternative divergence times and tree topologies

Despite having sampled many new individuals from islands worldwide, given the wide geographical scale of our study we still rely on sequence data for thousands of individuals submitted to GenBank over the years. Whenever multi-loci analyses including our focal taxa were available we used them; however, these are rare (Extended Data Table 4). Therefore, the majority of our phylogenies are based on a small number of genes, and most on a single gene, cytochrome *b*, which is the most widely sequenced mitochondrial marker in birds. Although some studies on island birds have shown that colonization and diversification times derived from mitochondrial trees often do not differ much from those obtained using multiple loci⁵⁷, it is possible

that in some cases the scaling and topologies of the trees might have been more accurate had we used multiple loci⁵⁸. This is particularly relevant for recent island colonists, given incomplete lineage sorting⁵⁹. An additional shortcoming of relying on published sequence data is that many of our DNA alignments often have substantial sections with missing data (for example, because only one small section of the gene could be sequenced and was uploaded to GenBank), which has been shown to lead to biases in branch lengths and topology⁶⁰. While future studies using phylogenomic approaches may address these issues, obtaining tissue samples for all of these taxa will remain an obstacle for a long time.

Although DAISIE does not directly use topological information (only divergence times are used), it is possible that the true topology for a clade may differ from that of the gene tree that we have estimated and this could have an effect on our results by (1) affecting colonization and branching times (addressed in the paragraph below); or (2) by altering the number of colonization events. Alternative topologies may have led to an increase or decrease in colonization events—for instance, some species that appear to have colonized an archipelago only once may have colonized multiple times and if these re-colonizations are recent they may go undetected when using one or few loci. As with any phylogenetic study, we cannot rule out this possibility, but we assume that recent re-colonization of the archipelagos in our dataset by the same taxon is rare, as these are all oceanic and isolated. For archipelago lineages with cladogenesis (26 out of 502 lineages), alternative topologies could include non-monophyly of island radiations, with the corollary being that they would be the result of multiple colonization events. However, this seems improbable for these isolated and well-studied radiations, for which morphological evidence (for example, HBW³⁷) is consistent with their monophyly as supported by existing molecular data.

Regarding scaling of divergence times, we assessed how uncertainty in our estimated node ages could influence our results by running an analysis of 100 datasets. For each dataset we sampled the node ages (that is, colonization and branching times) at random from a uniform distribution centred on the posterior mean for that node in the BEAST tree and extending twice the length of the highest posterior density (HPD) interval. For example, for a node with a 95% HPD interval of 2–3 million years in our trees, the uniform distribution was set to between 1.5 and 3.5 million years. The HPD interval will capture uncertainty under the selected phylogenetic and substitution models for the loci that we used, but we conduct our sensitivity analysis over a broader interval to accommodate the potential that the selected models and gene trees are inadequate. For cases in which using this approach meant that the lower bound of the uniform distribution was less than 0, we assigned a value of 0.00001 million years to the lower bound. We fitted the 9 best models to the 100 datasets using 5 initial starting parameters for each model (total 4,500 optimizations). We found that parameter estimates across the 100 datasets did not differ strongly from those in the main dataset (Supplementary Table 7). Notably, model selection was unaffected, with the M19 model being selected for all 100 datasets. This is because a lot of the information used for model selection is coming from the other sources of information that DAISIE uses (island age, number of species and endemism status) rather than colonization or branching times.

The maximum-likelihood parameters of the M19 model and the resulting area and isolation dependencies for datasets D1 to D6 (discussed below) are shown in Extended Data Fig. 6 and the DAISIE R objects including these alternative datasets are available in Mendeley Data (<https://doi.org/10.17632/sy58zbn3s2.2>).

To account for uncertainty in the rates of molecular evolution, we repeated all BEAST dating analyses for markers that were not cytochrome *b* using (1) the previously published cytochrome *b* rate⁵⁰ (dataset D1, equal to main dataset) and (2) previously estimated marker-specific rates⁴¹, which have also been widely used in the literature

(dataset D2). Although the trees dated using the marker-specific rates provide younger ages, we found that the DAISIE results were very similar using either approach (same model preferred and similar parameters). Therefore, in the main text we only discuss the results of analyses of D1, that is, applying the cytochrome *b* rate to all markers.

For some taxa, we did not use the stem age as the estimate of colonization time, and instead used alternative nodes (see ‘Colonization and branching times’). To test whether our choice of nodes affects our main conclusions, we recoded all such taxa by extracting the stem ages and used these ages as an upper bound for colonization (DAISIE MaxAge option). We fitted all 28 models to this new dataset (D3) and found that the M19 model is preferred and that the parameters and area or isolation relationships vary only slightly from those of the main analysis. We therefore conclude that our results are robust to the node selection approach.

If extinction has been high on the mainland, or if we failed to sample the closest relatives of the island taxa, this could lead to an overestimation of colonization times when using the stem age as the precise time of colonization. To investigate how this could have influenced our results, we ran analyses of datasets in which we allowed colonization to have happened at any time since the stem age (that is, the time of divergence from the nearest relative of the taxon on the mainland). For this we used the DAISIE options Endemic_MaxAge or NonEndemic_MaxAge, which integrate over all possible ages between the given maximum age and the present (or the first branching event within the archipelago for cases in which cladogenesis has occurred). We repeated this analysis coding all stem ages as maximum ages (D4), or coding only the 25% older stem ages as maximum ages (to account for the fact that older stems have the potential to have more bias) (D5). We also ran analyses on 100 datasets (D6) for which we assigned precise younger ages by randomly selecting a value between the stem age and the present (or crown age for cladogenetic groups). For all of these datasets (D4–D6), we found that the same model (M19) was preferred, but the initial values of the biogeographical rates (cladogenesis, extinction, colonization and anagenesis) were estimated to be higher than in the main dataset. Notably, the exponent hyperparameters were similar to those in the main dataset, meaning that the shape of the relationships between parameters and area or isolation is not much affected (Extended Data Fig. 6). The only exception is perhaps anagenesis, for which the relationships varied more markedly—with isolated islands achieving very high rates for this parameter—but still agreeing with our main conclusions. Anagenesis is in general the most difficult parameter to estimate⁵³. Thus, our conclusions are robust to the colonization times potentially being younger than those in our main dataset.

Sensitivity to archipelago selection and isolation metrics

The results of the following sensitivity analyses are presented in Supplementary Data 3 and the DAISIE R objects that include these alternative datasets are available in Mendeley Data (<https://doi.org/10.17632/sy58zbv3s2.2>).

To test whether the inclusion of both true archipelagos and single islands in our dataset could affect the results, we repeated analyses excluding single island units and found that the same model was preferred. The estimated initial rate of cladogenesis (λ_0) is higher if we exclude single islands, but this parameter is not different from a distribution of parameters estimated from datasets generated using a stratified-random sampling of both archipelagos and single islands.

Alternative isolation metrics to D_m have been shown to explain varying and often higher amounts of variation in species richness on islands⁶¹. We tested two alternative metrics: distance to the nearest larger or equivalent-sized landmass (D_b), and the mean between D_m and D_b (metrics given in Supplementary Data 2). We found that the same DAISIE model with very similar parameters was preferred in both cases, and we therefore used only the D_m metric, as this is more similar to the original model of MacArthur and Wilson.

The Mascarenes (Mauritius Island, Reunion and Rodrigues) are often treated as a single biogeographical unit in analyses. We chose to analyse them as independent units because (1) the distance between islands is much greater than our threshold for archipelago definition (more than 500 km between Mauritius Island and Rodrigues; more than 170 km between Reunion and Mauritius Island); (2) only two species of our target group are shared between the islands (*Terpsiphone bourbonnensis* is found in Mauritius Island and Reunion; and *Psittacula eques* is found in Mauritius Island and extirpated from Reunion), suggesting low connectivity; (3) although there are three clades whose branching events took place within the Mascarenes (*Coracina*, *Pezophaps* and *Raphus*, and *Zosterops*), the remaining species result from independent colonizations, suggesting that the three islands behave mostly as three different biogeographical units. We nevertheless ran an analysis treating the islands as a single archipelagic unit and found that the same model was preferred and with similar parameter estimates, and we therefore discuss only the results treating them as separate.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

New sequence data produced for this study have been deposited in GenBank with the accession codes: MH307408–MH307656. The following datasets have been deposited in Mendeley: DNA alignments (<https://doi.org/10.17632/vf95364vx6.1>), new phylogenetic trees produced for this study (<https://doi.org/10.17632/p6hm5w8s3b.2>), and DAISIE R objects (<https://doi.org/10.17632/sy58zbv3s2.2>). The 11 previously published trees are available upon request.

Code availability

The custom computer code used for this study is freely available in the DAISIE R package (<https://github.com/rsetienne/DAISIE>).

- Plummer, P. S. & Belle, E. R. Mesozoic tectono-stratigraphic evolution of the Seychelles microcontinent. *Sediment. Geol.* **96**, 73–91 (1995).
- Weigelt, P., Jetz, W. & Kreft, H. Bioclimatic and physical characterization of the world's islands. *Proc. Natl Acad. Sci. USA* **110**, 15307–15312 (2013).
- Norder, S. J. et al. Beyond the Last Glacial Maximum: island endemism is best explained by long-lasting archipelago configurations. *Glob. Ecol. Biogeogr.* **28**, 184–197 (2019).
- Thomson, J. & Walton, A. Redetermination of chronology of Aldabra atoll by ²³⁰Th/²³⁴U dating. *Nature* **240**, 145–146 (1972).
- Price, J. P. & Clague, D. A. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proc. R. Soc. B* **269**, 2429–2435 (2002).
- Valente, L. et al. Equilibrium bird species diversity in Atlantic islands. *Curr. Biol.* **27**, 1660–1666 (2017).
- del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & Kirwan, G. (eds.) *Handbook of the Birds of the World Alive* (Lynx Edicions, 2018).
- Valente, L., Etienne, R. S. & Dávalos, L. M. Recent extinctions disturb path to equilibrium diversity in Caribbean bats. *Nat. Ecol. Evol.* **1**, 0026 (2017).
- Steadman, D. W. *Extinction and Biogeography of Tropical Pacific Birds* (Univ. Chicago Press, 2006).
- Cheke, A. & Hume, J. P. *Lost Land of the Dodo: The Ecological History of Mauritius, Réunion and Rodrigues* (Bloomsbury, 2010).
- Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M. & Fleischer, R. C. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Curr. Biol.* **21**, 1838–1844 (2011).
- Rando, J. C., Pieper, H., Olson, S. L., Pereira, F. & Alcover, J. A. A new extinct species of large bullfinch (Aves: Fringillidae: *Pyrrhula*) from Graciosa Island (Azores, North Atlantic Ocean). *Zootaxa* **4282**, 567–583 (2017).
- Illera, J. C., Rando, J. C., Richardson, D. S. & Emerson, B. C. Age, origins and extinctions of the avifauna of Macaronesia: a synthesis of phylogenetic and fossil information. *Quat. Sci. Rev.* **50**, 14–22 (2012).
- Hume, J. P., Martill, D. & Hing, R. A terrestrial vertebrate palaeontological review of Aldabra atoll, Aldabra group, Seychelles. *PLoS ONE* **13**, e0192675 (2018).
- Cheke, A. S. Extinct birds of the Mascarenes and Seychelles—a review of the causes of extinction in the light of an important new publication on extinct birds. *Phelsuma* **21**, 4–19 (2013).
- Hume, J. P. & Walters, M. *Extinct Birds* (A&C Black, 2012).
- Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

48. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
49. Posada, D. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
50. Weir, J. T. & Schluter, D. Calibrating the avian molecular clock. *Mol. Ecol.* **17**, 2321–2328 (2008).
51. Field, D. J. et al. Timing the extant avian radiation: the rise of modern birds, and the importance of modeling molecular rate variation. *PeerJ Preprints* **7**, e27521v1 (2019).
52. Cicero, C. & Johnson, N. K. Higher-level phylogeny of new world Vireos (Aves: Vireonidae) based on sequences of multiple mitochondrial DNA genes. *Mol. Phylogenet. Evol.* **20**, 27–40 (2001).
53. Valente, L., Phillimore, A. B. & Etienne, R. S. Using molecular phylogenies in island biogeography: it's about time. *Ecography* **41**, 1684–1686 (2018).
54. Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816–1824 (2010).
55. Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B* **344**, 305–311 (1994).
56. Etienne, R. S. et al. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B* **279**, 1300–1309 (2012).
57. Stervander, M. et al. Disentangling the complex evolutionary history of the Western Palearctic blue tits (*Cyanistes* spp.) — phylogenomic analyses suggest radiation by multiple colonization events and subsequent isolation. *Mol. Ecol.* **24**, 2477–2494 (2015).
58. Ogilvie, H. A., Heled, J., Xie, D. & Drummond, A. J. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. *Syst. Biol.* **65**, 381–396 (2016).
59. Maddison, W. P. & Knowles, L. L. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* **55**, 21–30 (2006).
60. Lemmon, A. R., Brown, J. M., Stanger-Hall, K. & Lemmon, E. M. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* **58**, 130–145 (2009).
61. Weigelt, P. & Kreft, H. Quantifying island isolation—insights from global patterns of insular plant species richness. *Ecography* **36**, 417–429 (2013).
62. Nielson, D. L. & Sibbett, B. S. Geology of Ascension Island, South Atlantic Ocean. *Geothermics* **25**, 427–448 (1996).
63. Ramalho, R. S. et al. Emergence and evolution of Santa Maria Island (Azores)—the conundrum of uplifted islands revisited. *Geol. Soc. Am. Bull.* **129**, 372–390 (2017).
64. Hearty, P. J. & Olson, S. L. Geochronology, biostratigraphy, and changing shell morphology in the land snail subgenus *Poecilozonites* during the Quaternary of Bermuda. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **293**, 9–29 (2010).
65. Carracedo, J. C. & Troll, V. R. *The Geology of the Canary Islands* (Elsevier, 2016).
66. Ramalho, R. *Building the Cape Verde Islands* (Springer, 2011).
67. Eisenhauer, A., Heiss, G. A., Sheppard, C. R. C. & Dullo, W. C. in *Ecology of the Chagos Archipelago* (eds. Sheppard, C. R. C. & Seaward, M. R. D.) 21–31 (Linnean Society Occasional Publications, 1999).
68. Campbell, H. J. Fauna and flora of the Chatham Islands: less than 4 m.y. old. *Geological Society of New Zealand Miscellaneous Publication* 97 (eds Cooper, R. A. & Jones, C.) 15–16 (Geological Society of New Zealand, 1998).
69. Bullough, F. History and geology of Christmas Island. *Geological Society of London Blog* <https://blog.geolsoc.org.uk/2013/12/18/door-18-history-and-geology-of-christmas-island/> (2013).
70. Castillo, P. et al. Anomalously young volcanoes on old hot-spot traces: I. Geology and petrology of Cocos Island. *Geol. Soc. Am. Bull.* **100**, 1400–1414 (1988).
71. Woodroffe, C. D., Veeh, H. H., Falkland, A. C., McLean, R. F. & Wallensky, E. Last interglacial reef and subsidence of the Cocos (Keeling) Islands, Indian Ocean. *Mar. Geol.* **96**, 137–143 (1991).
72. Nougier, J., Cantagrel, J. M. & Karche, J. P. The Comores archipelago in the western Indian Ocean: volcanology, geochronology and geodynamic setting. *J. Afr. Earth Sci.* **5**, 135–144 (1986).
73. Almeida, F. in *Sítios Geológicos e Paleontológicos do Brasil* (eds Schobbenhaus, C. et al.) 361–368 (Comissão Brasileira de Sítios Geológicos e Paleobiológicos, 2000).
74. Ali, J. R. & Aitchison, J. C. Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *J. Biogeogr.* **41**, 1227–1241 (2014).
75. Ryan, P. G. in *Encyclopedia of Islands* (eds Gillespie, R. & Clague, D.) 929–932 (Univ. California Press, 2009).
76. Batiza, R. Petrology and chemistry of Guadalupe Island: an alkalic seamount on a fossil ridge crest. *Geology* **5**, 760–764 (1977).
77. Stuessy, T. F., Foland, K. A., Sutter, J. F., Sanders, R. W. & Silva O., M. Botanical and geological significance of potassium–argon dates from the Juan Fernandez islands. *Science* **225**, 49–51 (1984).
78. McDougall, I., Embleton, B. J. J. & Stone, D. B. Origin and evolution of Lord Howe Island, Southwest Pacific Ocean. *J. Geol. Soc. Aust.* **28**, 155–176 (1981).
79. Mata, J. et al. in *Geologia de Portugal, Volume II Geologia Meso-cenozoica de Portugal* (eds Dias, R. et al.) 691–746 (Escolar Editora, 2013).
80. Guille, G. et al. Les marquises (Polynésie Françaises): un archipel intraocéanique atypique. *Geol. France* **2**, 5–36 (2002).
81. Montaggioni, L. & Nativel, P. *La Reunion, Ile Maurice. Géologie et Aperçus Biologiques, Plantes et Animaux* (Masson, 1988).
82. Grandcolas, P. et al. New Caledonia: a very old Darwinian island? *Phil. Trans. R. Soc. B* **363**, 3309–3317 (2008).
83. Anthoni, J. Geography and geology of Niue. <http://www.seafriends.org.nz/niue/geo.htm> (2005).
84. Jones, J. G. & McDougall, I. Geological history of Norfolk and Philip islands, southwest Pacific Ocean. *J. Geol. Soc. Aust.* **20**, 239–254 (1973).
85. Suzuki, M., Taisuke, S. & Hideo, T. *Nomination of the Ogasawara Islands for Inscription on the World Heritage List* (Government of Japan, 2010).
86. Neall, V. E. & Trewick, S. A. The age and origin of the Pacific islands: a geological overview. *Phil. Trans. R. Soc. B* **363**, 3293–3308 (2008).
87. Hekinian, R. et al. The Pitcairn hotspot in the South Pacific: distribution and composition of submarine volcanic sequences. *J. Volcanol. Geotherm. Res.* **121**, 219–245 (2003).
88. Vezzoli, L. & Acocella, V. Easter Island, SE Pacific: an end-member type of hotspot volcanism. *Bull. Geol. Soc. Am.* **121**, 869–886 (2009).
89. Gillot, P.-Y., Lefèvre, J.-C. & Nativel, P.-E. Model for the structural evolution of the volcanoes of Réunion Island. *Earth Planet. Sci. Lett.* **122**, 291–302 (1994).
90. Safford, R. & Hawkins, F. *The Birds of Africa: Volume VIII: The Malagasy Region: Madagascar, Seychelles, Comoros, Mascarenes* (A&C Black, 2013).
91. Baker, I., Gale, N. H. & Simons, J. Geochronology of the St Helena volcanoes. *Nature* **215**, 1451–1456 (1967).
92. Duncan, R. A. in *Investigations of the Northern Melanesian Borderland, Earth Science Series* (Circum Pacific Council Publications, 1985).
93. Lee, D. C., Halliday, A. N., Fittin, J. G. & Poli, G. Isotopic variations with distance and time in the volcanic islands of the Cameroon line: evidence for a mantle plume origin. *Earth Planet. Sci. Lett.* **123**, 119–138 (1994).
94. Geldmacher, J., Hoernle, K., Van Den Bogaard, P., Zankl, G. & Garbe-Schönberg, D. Earlier history of the >70-Ma-old Canary hotspot based on the temporal and geochemical evolution of the Selvagen Archipelago and neighboring seamounts in the Eastern North Atlantic. *J. Volcanol. Geotherm. Res.* **111**, 55–87 (2001).
95. Clouard, V. & Bonneville, A. A in *Plates, Plumes and Paradigms* (eds Foulger, G. R. et al.) 71–90 (Geological Society of America, 2005).
96. Bohron, W. A. et al. Prolonged history of silicic peralkaline volcanism in the eastern Pacific Ocean. *J. Geophys. Res.* **101**, 11457–11474 (1996).
97. Kroenke, L. W. in *The Origin and Evolution of Pacific Island Biotas, New Guinea to Eastern Polynesia: Patterns and Processes* (eds. Keast, A. & Miller, S.) 19–34 (SPD Academic Publishing, 1996).
98. Ollier, C. D. Geomorphology of South Atlantic volcanic islands. Part I: the Tristan da Cunha group. *Z. Geomorphol.* **28**, 367–382 (1984).
99. Kocher, T. D. et al. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl Acad. Sci. USA* **86**, 6196–6200 (1989).
100. Dietzen, C., Witt, H.-H. & Wink, M. The phylogeographic differentiation of the European robin *Erithacus rubecula* on the Canary Islands revealed by mitochondrial DNA sequence data and morphometrics: evidence for a new robin taxon on Gran Canaria? *Avian Sci.* **3**, 115–132 (2003).
101. Edwards, S. V., Arctander, P. & Wilson, A. C. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. R. Soc. Lond. B* **243**, 99–107 (1991).
102. Helm-Bychowski, K. & Cracraft, J. Recovering phylogenetic signal from DNA sequences: relationships within the corvine assemblage (class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-*b* gene. *Mol. Biol. Evol.* **10**, 1196–1214 (1993).
103. Warren, B. H., Bermingham, E., Bowie, R. C. K., Prys-Jones, R. P. & Thébaud, C. Molecular phylogeography reveals island colonization history and diversification of western Indian Ocean sunbirds (*Nectarinia*: Nectariniidae). *Mol. Phylogenet. Evol.* **29**, 67–85 (2003).
104. Farrington, H. L., Lawson, L. P., Clark, C. M. & Petren, K. The evolutionary history of Darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. *Evolution* **68**, 2932–2944 (2014).
105. Warren, B. H. et al. Hybridization and barriers to gene flow in an island bird radiation. *Evolution* **66**, 1490–1505 (2012).
106. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thébaud, C. Tracking island colonization history and phenotypic shifts in Indian Ocean bulbuls (*Hypsipetes*: Pycnonotidae). *Biol. J. Linn. Soc.* **85**, 271–287 (2005).
107. Andersen, M. J., Hosner, P. A., Filardi, C. E. & Moyle, R. G. Phylogeny of the monarch flycatchers reveals extensive paraphyly and novel relationships within a major Australo-Pacific radiation. *Mol. Phylogenet. Evol.* **83**, 118–136 (2015).
108. Sari, E. H. R. & Parker, P. G. Understanding the colonization history of the Galápagos flycatcher (*Myiarchus magnirostris*). *Mol. Phylogenet. Evol.* **63**, 244–254 (2012).
109. Chaves, J. A., Parker, P. G. & Smith, T. B. Origin and population history of a recent colonizer, the yellow warbler in Galápagos and Cocos Islands. *J. Evol. Biol.* **25**, 509–521 (2012).
110. Martínez-Gómez, J. E., Barber, B. R. & Peterson, A. T. Phylogenetic position and generic placement of the Socorro wren (*Thryomanes sissonii*). *Auk* **122**, 50–56 (2005).
111. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thébaud, C. Immigration, species radiation and extinction in a highly diverse songbird lineage: white-eyes on Indian Ocean islands. *Mol. Ecol.* **15**, 3769–3786 (2006).
112. McGuire, J. A. et al. Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.* **24**, 910–916 (2014).
113. Derryberry, E. P. et al. Lineage diversification and morphological evolution in a large-scale continental radiation: the neotropical ovenbirds and woodcreepers (Aves: Furnariidae). *Evolution* **65**, 2973–2986 (2011).
114. Jønsson, K. A. et al. A supermatrix phylogeny of corvid passerine birds (Aves: Corvidae). *Mol. Phylogenet. Evol.* **94**, 87–94 (2016).
115. Scofield, R. P. et al. The origin and phylogenetic relationships of the New Zealand ravens. *Mol. Phylogenet. Evol.* **106**, 136–143 (2017).
116. Cibois, A., Thibault, J. C., Bonillo, C., Filardi, C. E. & Pasquet, E. Phylogeny and biogeography of the imperial pigeons (Aves: Columbidae) in the Pacific Ocean. *Mol. Phylogenet. Evol.* **110**, 19–26 (2017).
117. Friis, G., Aleixandre, P., Rodríguez-Estrella, R., Navarro-Sigüenza, A. G. & Milá, B. Rapid postglacial diversification and long-term stasis within the songbird genus *Junco*: phylogeographic and phylogenomic evidence. *Mol. Ecol.* **25**, 6175–6195 (2016).
118. Marki, P. Z. et al. Supermatrix phylogeny and biogeography of the Australasian Meliphagidae radiation (Aves: Passeriformes). *Mol. Phylogenet. Evol.* **107**, 516–529 (2017).
119. Fuchs, J. et al. Long-distance dispersal and inter-island colonization across the western Malagasy Region explain diversification in brush-warblers (Passeriformes: *Nesillas*). *Biol. J. Linn. Soc.* **119**, 873–889 (2016).

120. Cibois, A. et al. Phylogeny and biogeography of the fruit doves (Aves: Columbidae). *Mol. Phylogenet. Evol.* **70**, 442–453 (2014).
121. Carmi, O., Witt, C. C., Jaramillo, A. & Dumbacher, J. P. Phylogeography of the vermilion flycatcher species complex: multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. *Mol. Phylogenet. Evol.* **102**, 152–173 (2016).
122. Cornetti, L. et al. The genome of the ‘great speciator’ provides insights into bird diversification. *Genome Biol. Evol.* **7**, 2680–2691 (2015).

Acknowledgements We thank the skilled guides and field assistants who helped with sample collection in the field; the ornithologists and collection curators who were kind enough to reply to requests for material; T. von Rintelen, K. von Rintelen and C. Zorn for support or advice; A. Pigot for comments on the manuscript; N. Bunbury (Seychelles Islands Foundation), who organized sample loans of Aldabra island; J. van de Crommenacke, J. Groombridge and H. Jackson for providing samples or DNA sequences; J. A. Alcover, J. C. Rando, F. Sayol and S. Faurby for sharing data on extinct species; C. Baeta, M. Hammers, J. Hume, D. Shapiro, J. Varela and P. Cascão for permission to use photographs or illustrations; P. Hearty, R. Stern and M. Reagan for expertise on island geological ages; A. Cibois, J. McGuire, H. Lerner, P. Marki, B. Milá, G. Friis, J. Fuchs, J. P. Dumbacher and O. Carmi for providing phylogenetic data; P. Weigelt for map data. We thank the following for permission to obtain new samples or to access existing samples, and for logistic support (locations are given in brackets): A. Carvalho and the Department of the Environment (São Tomé and Príncipe); J. Obiang, N. Calvo and the Universidad Nacional de Guinea Ecuatorial for Bioko and Annobón samples (Equatorial Guinea); the Ministry of Environment, Energy and Climate Change of the Republic of Seychelles, the Seychelles Bureau of Standards, BirdLife Seychelles and Seychelles Islands Foundation (Seychelles); Centre National de Documentation et de Recherche Scientifique (Grande Comore & Anjouan), Action Comores, Direction de l’Agriculture et de la Forêt (Mayotte) (Comoros); Ministère des Eaux et Forêts (Madagascar) and the Madagascar Institute pour la Conservation des Ecosystemes Tropicaux (Madagascar); Mauritius National Parks and Conservation Service and Mauritius Wildlife Foundation (Mauritius); O. Hébert, W. Waheoneme, N. Clark, the Direction de L’Environnement (South Province), Direction du Développement Economique (Loyalty Islands Province), and local chiefs and landowners (New Caledonia); Moroccan Environment Ministry (Morocco); Cape Verde Agriculture and Environment Ministry (Cape Verde); F. Njie and the Limbe Botanical and Zoological Garden (Cameroon); Station de Recherche de

l’IRET at Ipassa-Makokou (Gabon); Fernanda Lages (ISCED-Huíla) (Angola); the regional governments of Andalucía and the Canary Islands (Spain); regional governments of Madeira and the Azores (Portugal). We thank the Department of Ornithology and Mammalogy of the California Academy of Sciences (L. Wilkinson and M. Flannery) for loaning Galápagos samples; the Natural History Museum at Tring (M. Adams) for loaning Comoros samples; the Stuttgart State Museum of Natural History for loaning stonechat samples from Madagascar. S. Block assisted with cluster analyses at the Museum für Naturkunde. The Center for Information Technology of the University of Groningen provided support and access to the Peregrine high-performance computing cluster. L.V. was funded by the German Science Foundation (DFG Research grant VA 1102/1-1), the Alexander von Humboldt Foundation, the Brandenburg Postdoc Prize 2015 and by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO); R.S.E. was supported by a NWO VICI grant; M.M. was supported by the Portuguese Science and Technology Foundation (post-doctoral grant: SFRH/BPD/100614/2014); S.M.C. was supported by the National Geographic Society (CRE grant 9383-13); J.C.I. was supported by the Spanish Ministry of Science, Innovation and Universities (PGC2018-097575-B-I00) and by a GRUPIN research grant from the Regional Government of Asturias (IDI/2018/000151); and C.T. was supported by the ‘Laboratoire d’Excellence’ TULIP (ANR-10-LABX-41).

Author contributions L.V., A.B.P. and R.S.E. designed the study, developed the analytical framework and performed statistical analyses. L.V. compiled the data, conducted most of the analyses and wrote the first draft. R.S.E. developed the likelihood method. A.B.P. and R.S.E. contributed substantially to the writing. M.M., B.H.W., S.M.C., J.C.I. and C.T. provided expertise on island birds, collected bird tissue samples, and provided molecular and/or phylogenetic data. K.H. and J.C.I. performed laboratory work. R.T. contributed to molecular analyses. T.A. performed analyses. All authors commented on the draft.

Competing interests The authors declare no competing interests.

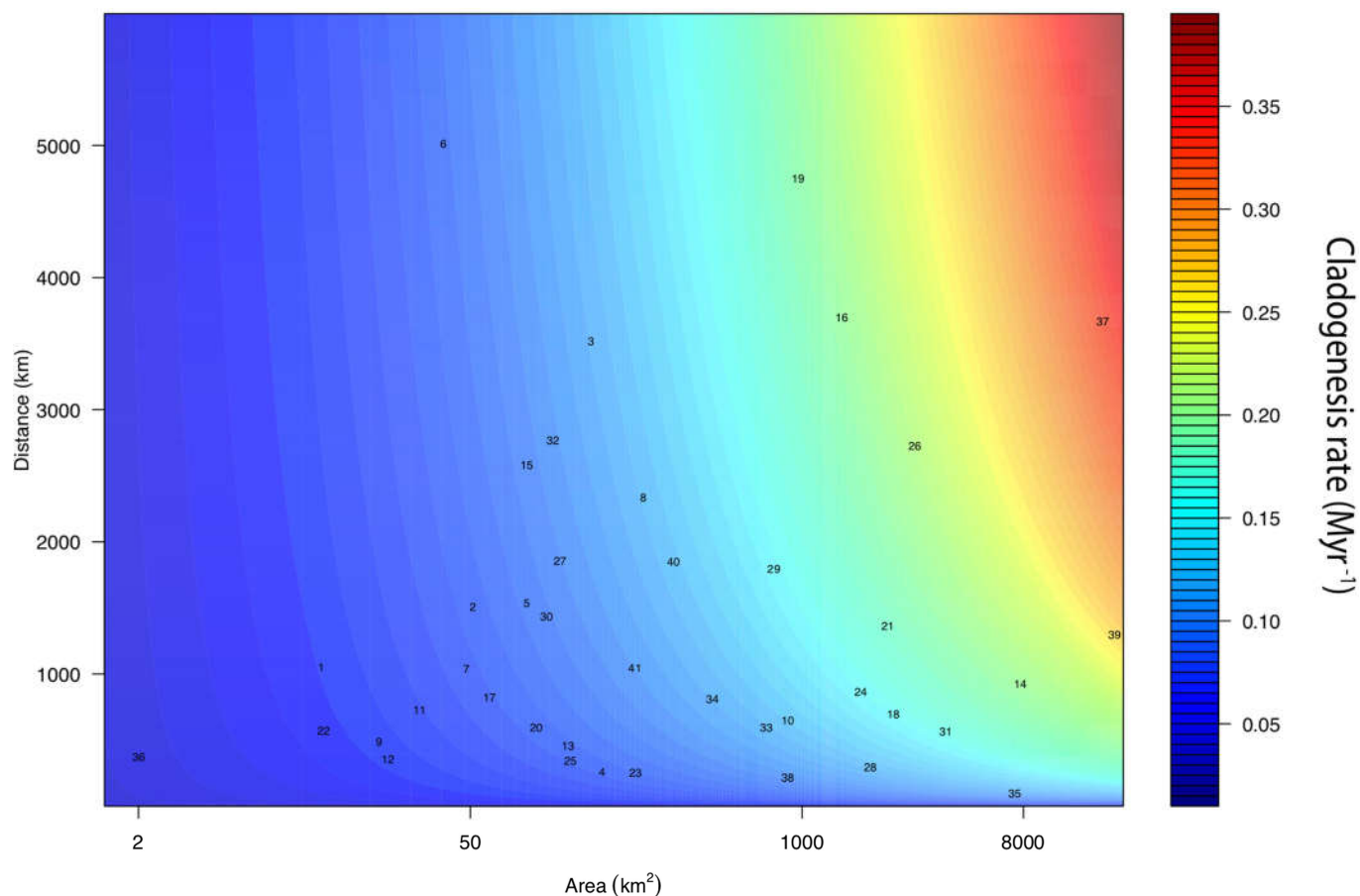
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2022-5>.

Correspondence and requests for materials should be addressed to L.V.

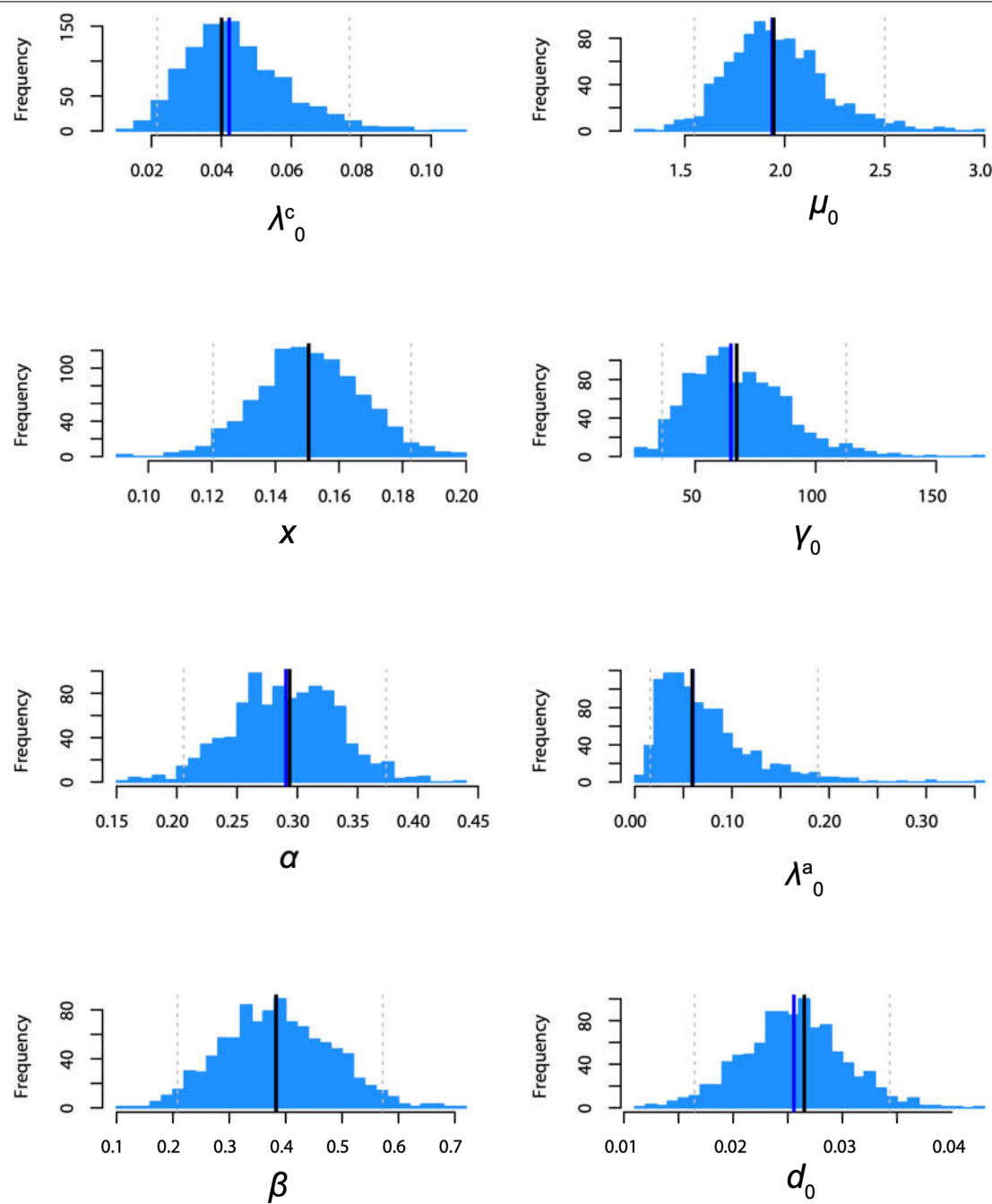
Peer review information *Nature* thanks Thomas Matthews, Kostas Triantis, Jason Weir and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



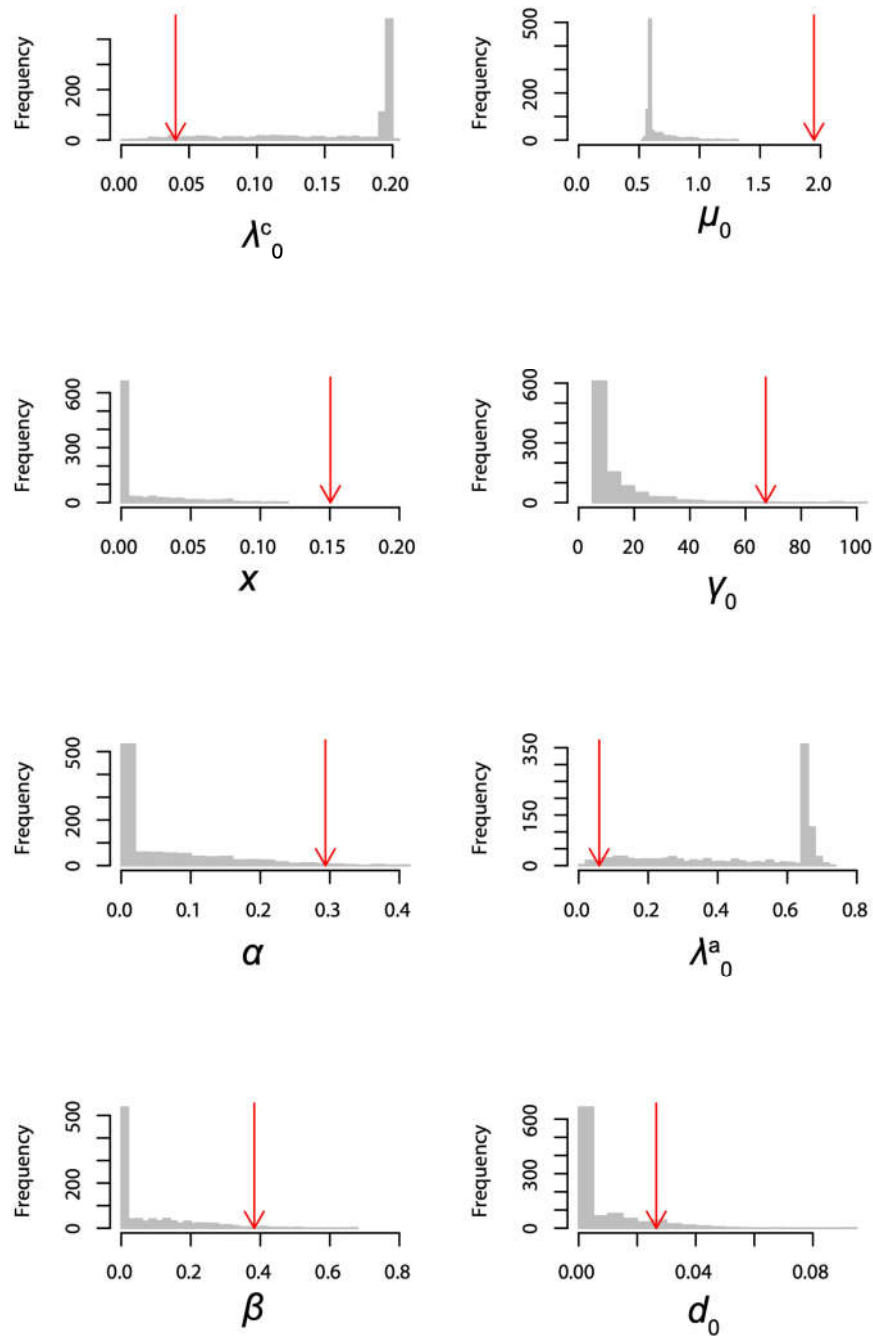
Extended Data Fig. 1 | Variation of cladogenesis with isolation and area. Contour plot showing how the local rate of cladogenesis varies with area and D_m assuming the maximum-likelihood global hyperparameters of the M19 model (equations describing the relationships are provided in Supplementary

Table 1). Numbers correspond to the archipelago numbers from Fig. 1 and show the local cladogenesis rates for each of the archipelagos in our dataset. Area is shown as a log scale.



Extended Data Fig. 2 | Bootstrap precision estimates of the parameters of the M19 model. Parametric bootstrap analysis fitting the M19 model to 1,000 global datasets simulated with maximum-likelihood parameters of the M19 model. Plots are frequency histograms of estimated parameters. Black

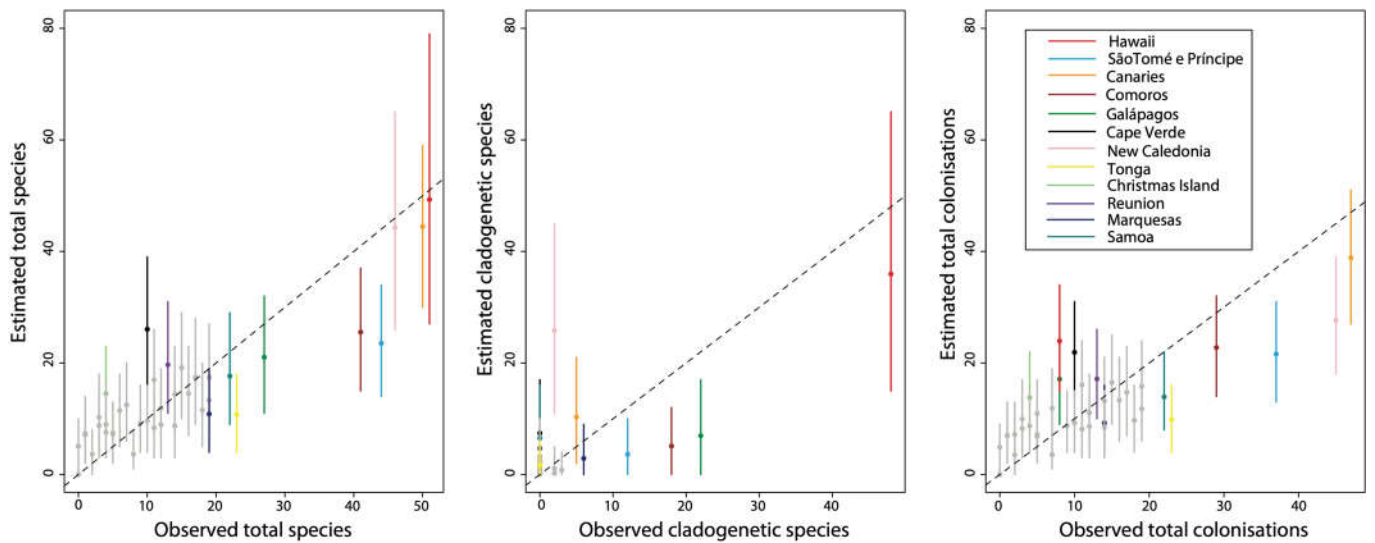
lines show the median estimated values across all simulations and the blue lines the simulated values. Dashed lines show 2.5–97.5 percentiles. Parameters are explained in Supplementary Table 1. Bootstrap parameter estimates for the M14 model are shown in Extended Data Table 5.



Extended Data Fig. 3 | Randomization analysis of the M19 model.

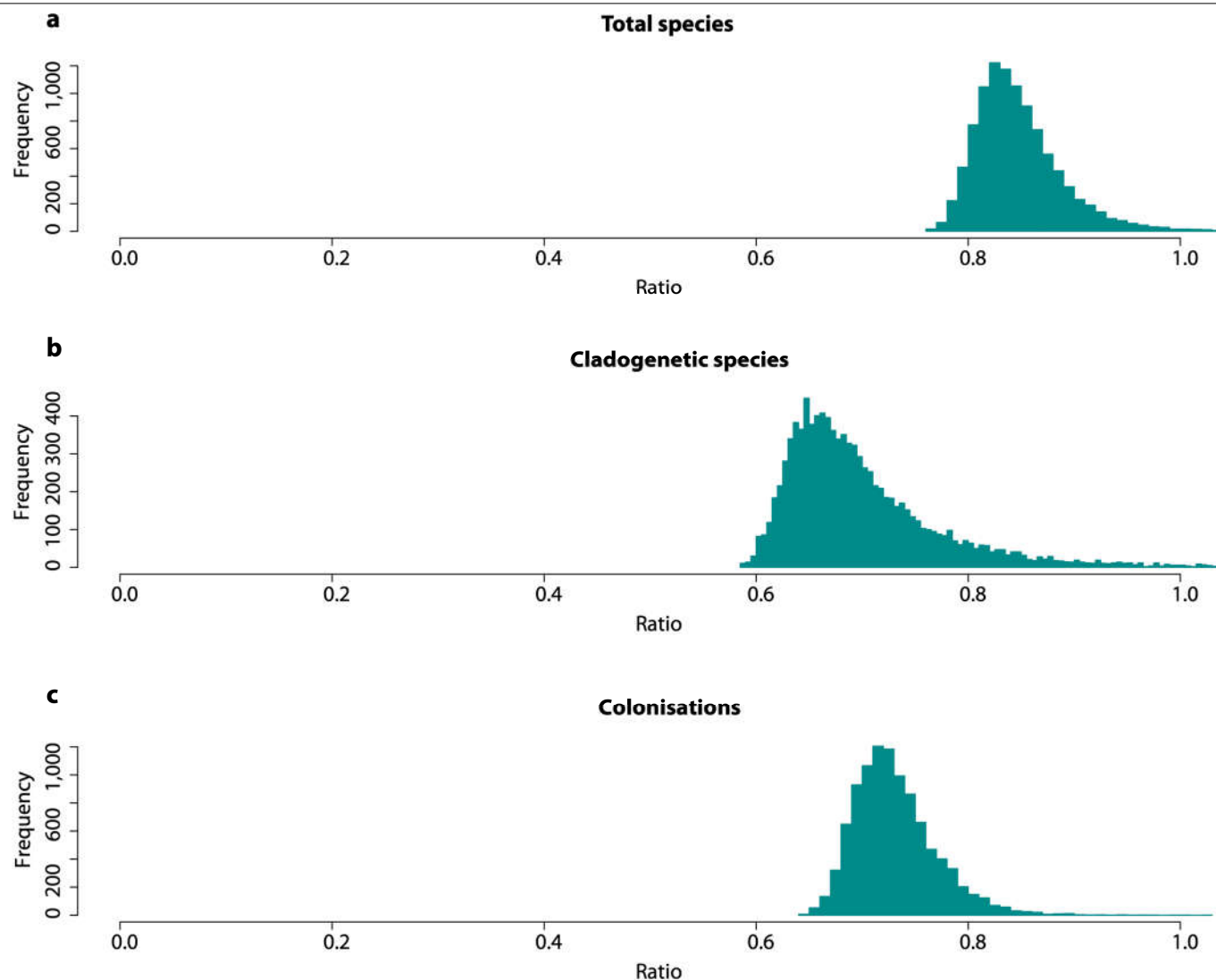
Distribution of global hyperparameters estimated from each of 1,000 datasets with the same phylogenetic data as our main global dataset but randomly reshuffling archipelago area and isolation among the 41 archipelagos in the dataset. Grey histograms show DAISIE maximum-likelihood parameter

estimates for the M19 model. Red arrows show the estimated parameter from the real data. In most cases, the hyperparameters describing the exponent of the power models (x , α , β and d_0) are estimated as zero in the reshuffled datasets, which is not the case in the real data (red). Parameters are explained in Supplementary Table 1.



Extended Data Fig. 4 | Goodness of fit of the preferred model (M19). Plots show the observed total number of species, cladogenetic species and colonisations versus those simulated by the model. Median and 95% percentiles are shown for 1,000 simulations of each archipelago. Selected

archipelagos mentioned in the main text or well-known archipelagos for which one or more of the diversity metrics are under- or overestimated are highlighted in colour. Dashed line is $y=x$. See also Fig. 3.



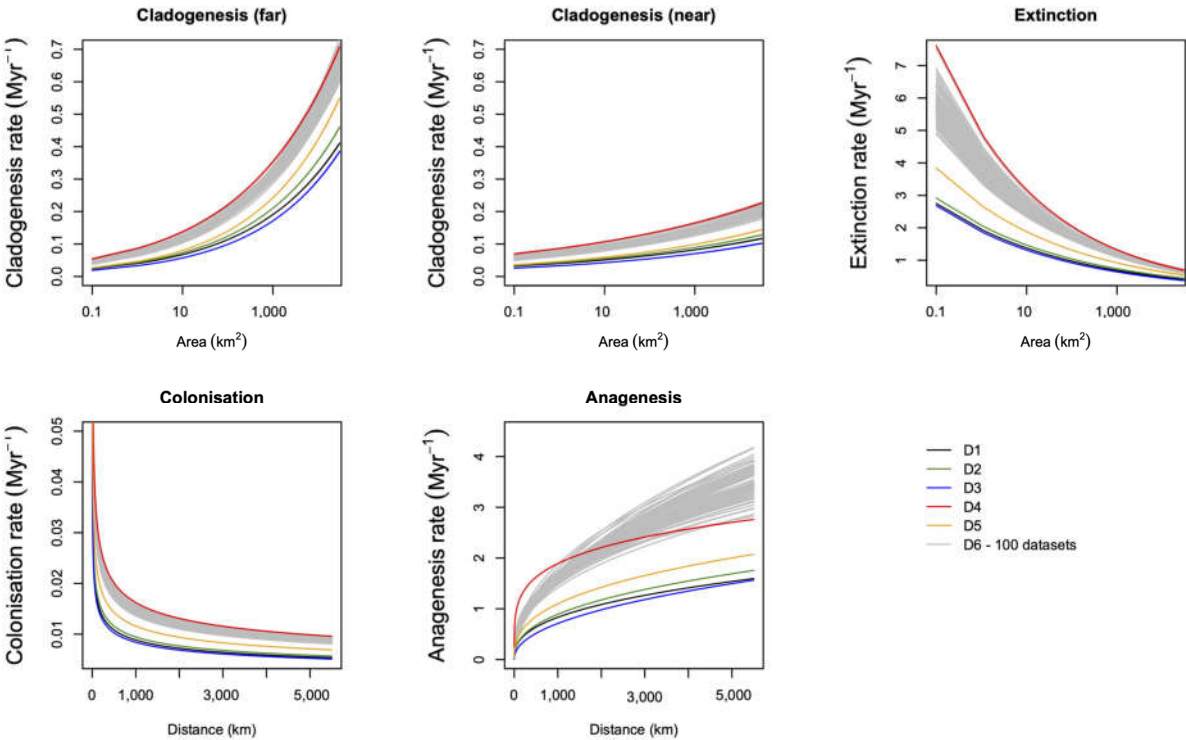
Extended Data Fig. 5 | Ratio of pseudo- R^2 observed over pseudo- R^2 simulated. Estimates were based on 10,000 datasets simulated using the M19 model. A ratio centred on 1 would indicate that the model explains the

observed data as well as it is able to explain the average dataset simulated under the maximum-likelihood parameters.

a

Dataset	Characteristics	Cladogenesis			Extinction		Colonisation		Anagenesis	
		λ_0	γ	d_0	μ_0	x	γ_0	α	λ^a_0	β
D1	Main dataset. <i>cyt-b</i> molecular rate applied to all markers.	0.040	0	0.027	1.946	0.150	67.256	0.294	0.059	0.383
D2	Molecular rates from Lerner et al. applied to non- <i>cyt-b</i> markers	0.043	0	0.027	2.070	0.149	73.460	0.297	0.057	0.398
D3	Same as D1 but choosing stem ages for all colonisation times	0.033	0	0.028	1.877	0.155	63.106	0.292	0.029	0.463
D4	Same as D1 but using all ages as a maximum age	0.086	0	0.024	4.914	0.190	137.196	0.309	0.408	0.222
D5	Same as D1 but considering the 25% oldest stem ages in the dataset as maximum ages	0.046	0	0.028	2.687	0.155	94.449	0.304	0.083	0.373
D6	Same as D1 but with all ages younger (100 datasets with random younger ages)	0.06 - 0.08	0	0.02 - 0.03	3.47 - 4.48	0.16 - 0.19	113.15 - 132	0.30 - 0.31	0.04 - 0.12	0.38 - 0.52

b



Extended Data Fig. 6 | Sensitivity to colonization and branching times.
a, Maximum-likelihood parameter estimates of the M19 model (preferred model) for datasets differing in colonization and branching times. D6 represents 100 datasets, therefore, the 2.5 and 97.5 percentiles are shown. Parameter symbols are described in Supplementary Table 1. **b**, Estimated

relationships between island area and isolation and local island biogeography parameters for each dataset. Under the M19 model, cladogenesis rate increases with both area and isolation, and thus plots for more (far, 5,000 km) and less (near, 50 km) isolated islands are shown.

Extended Data Table 1 | Archipelago characteristics and references for island geological ages

Archipelago	Area (km ²)	Distance nearest mainland (km)	Age (Ma)	Reference age	Number of species in focal group	Species sampled in phylogenies (percentage)	Colonisations sampled in phylogenies (percentage)
Aldabra Group	180	261	1	*	12	11 (91.67)	11 (91.67)
Ascension	91	1537	1	⁶²	0	NA (NA)	NA (NA)
Azores	2387	1365	6.3	⁶³	17	13 (76.47)	13 (76.47)
Bermuda	53	1040	2	^{64†}	5	2 (40)	2 (40)
Canary Islands	7493	96	21	⁶⁵	50	44 (88)	41 (87.23)
Cape Verde	4033	570	15.8	⁶⁶	10	10 (100)	10 (100)
Chagos	56.13	1510	0.0065	⁶⁷	0	NA (NA)	NA (NA)
Chatham	966	650	3	⁶⁸	14	13 (92.86)	13 (92.86)
Christmas Island	135	345	10	⁶⁹	4	2 (50)	2 (50)
Cocos (Costa Rica)	24	491	2.4	⁷⁰	4	2 (50)	2 (50)
Cocos (Keeling)	14.2	1054	0.003	⁷¹	0	NA (NA)	NA (NA)
Comoros	2033	297	15	⁷²	41	40 (97.56)	28 (96.55)
Fernando de Noronha	26	360	3.3	⁷³	3	1 (33.33)	1 (33.33)
Galápagos	7880	928	4	⁷⁴	27	26 (96.3)	8 (100)
Gough	91	2582	4	⁷⁵	1	1 (100)	1 (100)
Guadalupe	244	254	7	⁷⁶	11	10 (90.91)	10 (90.91)
Hawaii	16624.6	3670	29.8	³⁵	51	31 (60.78)	6 (75)
Juan Fernández	99.67	600	5.8	⁷⁷	6	6 (100)	5 (100)
Lord Howe	14.55	571	6.9	⁷⁸	11	9 (81.82)	9 (81.82)
Madeira	798	600	18.8	⁷⁹	19	17 (89.47)	17 (89.47)
Marianas (with Guam)	852	1800	15	‡	19	11 (57.89)	11 (57.89)
Marquesas	1063	4750	5.5	⁸⁰	19	12 (63.16)	7 (50)
Mauritius Isl.	1865	867	8.9	⁸¹	15	9 (60)	9 (60)
New Caledonia	18576	1300	37	⁸²	46	38 (82.61)	37 (82.22)
Niue	261.46	2340	2	⁸³	5	3 (60)	3 (60)
Norfolk	34.6	730	3.05	⁸⁴	14	11 (78.57)	11 (78.57)
Ogasawara	65	827	5	⁸⁵	10	5 (50)	5 (50)
Palau	488	815	20.1	⁸⁶	16	10 (62.5)	10 (62.5)
Pitcairn	42.8	5015	1.1	⁸⁷	8	3 (37.5)	2 (28.57)
Rapa Nui	163.6	3519	0.78	⁸⁸	2	0 (0)	0 (0)
Reunion	2512	700	5	⁸⁹	13	9 (69.23)	9 (69.23)
Rodrigues	109	1440	15	⁹⁰	9	4 (44.44)	4 (44.44)
Saint Helena	123.28	1856	14.5	⁹¹	3	0 (0)	0 (0)
Samoa	3041	2730	13.5	⁹²	22	16 (72.73)	16 (72.73)
São Tomé and Príncipe	964	219	30	⁹³	44	44 (100)	37 (100)
Selvagens	2.73	373	29	⁹⁴	1	1 (100)	1 (100)
Seychelles Inner	242.68	1048	64	³¹	12	11 (91.67)	11 (91.67)
Society Islands	1577.8	3700	4.3	⁹⁵	18	7 (38.89)	7 (38.89)
Socorro	132	457	3.5	⁹⁶	7	7 (100)	7 (100)
Tonga	344.4	1850	41	⁹⁷	23	10 (43.48)	10 (43.48)
Tristan da Cunha	115.4	2770	18	⁹⁸	4	4 (100)	2 (100)

Island ages are from previously published studies^{31,35,62–98}.

More data are provided in Supplementary Data 2. For archipelagos closer to Madagascar, New Guinea or New Zealand than to the continent, we use those islands as the mainland.

*A previous study³⁴ proposed an age of 0.125 million years, but we used an older age (see Methods).

†At least 2 million years (P. Hearty, personal communication).

‡R. Stern and M. K. Reagan, personal communication.

Extended Data Table 2 | Primer sequences used in this study

Primer name	Sequence	Reference
New primers designed		
L-cytB_bird_0	TCAACRACTCCCTAATYGACCT	This paper
H-cytB_bird_2	AGRAYTACTCCTGTGTTTCARGTYTC	This paper
L-cytB_bird_2	GARACYTGAAACACAGGAGTARTYCT	This paper
H-cytB_bird_3	TAGKGGGTTGTTTGAGCCTGWTTCTGTG	This paper
L-cytB_bird_3	CACGAAWCAGGCTCAAACAACCC	This paper
H-cytB_bird_4	GGAGTAGTADGGGTGAAATGGRATTTT	This paper
H-cytB_bird_5	GGGTGTTCTACTGGTTGGCTKCC	This paper
L-cytB_bird_5	CCMCTCTCACAAAYCCTATTCTGA	This paper
L-cytB_human	TGAAACTTCGGATCCCTACTA	This paper
Published primers		
L14841	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA	99
L14995	GCCCCATCCAACATCTCAGCATGATGAAACTCCG	100
L15308	GGC TAT GTC CTC CCA TGA GGC CAA AT	101
H15767	ATGAAGGGATGTTCTACTGGTTG	101
H15917	TAGTTGGCCAATGATGATGAATGGGTGTTCTACTGGTT	100
H16065	GAGTCTTCAGTCTCTGGTTTACAAGAC	102
L-cytB_Passer	CACAGGCCTAATTAAGCCTACCT	36
H-cytB_Passer	TTGARAATGCCAGCTTTGGGAG	36
L-cytB-Mot	CCAAATYGTTACAGGMCTCCTG	36
H-cytB-Mot	GGTGAATGAGGCTAGTTGCCCA	36

Primer sequences were designed for this study or are from previously published studies^{36,99–102}.

Extended Data Table 3 | The 80 alignments used in the phylogenetic analyses

Taxonomic group	Molecular marker(s)	Main source of sequences	Taxonomic group	Molecular marker(s)	Main source of sequences
<i>Acrocephalus</i>	cyt- <i>b</i>	-	<i>Moho</i>	cyt- <i>b</i>	-
Alaudidae (family)	cyt- <i>b</i>	-	<i>Monarcha</i>	cyt- <i>b</i>	-
<i>Alopecoenas</i> / <i>Gallicolumba</i>	ND2	-	<i>Motacilla</i>	cyt- <i>b</i>	-
<i>Anairetes</i>	cyt- <i>b</i>	-	<i>Myadestes</i>	cyt- <i>b</i>	-
<i>Anthus</i>	cyt- <i>b</i>	-	<i>Myiagra</i>	cyt- <i>b</i> + ND2	107
<i>Aphrastura</i>	COI	-	<i>Myiarchus</i>	cyt- <i>b</i> + ND2	108
<i>Bucanetes</i>	cyt- <i>b</i>	-	<i>Nigrita</i>	cyt- <i>b</i>	-
Buntings <i>Nesospiza</i> / <i>Rowettia</i>	cyt- <i>b</i>	-	<i>Onychognathus</i>	ND2	-
<i>Carduelis</i>	cyt- <i>b</i>	-	<i>Passer</i> / <i>Petronia</i>	cyt- <i>b</i>	-
<i>Chasiempis</i> / <i>Elepaio</i>	ND2	-	<i>Petroica</i>	cyt- <i>b</i>	-
<i>Chaunoproctus</i>	ND2	-	<i>Phylloscopus</i>	cyt- <i>b</i>	-
<i>Cinnyris notata</i>	ATP6	103	<i>Pipilo</i>	cyt- <i>b</i>	-
<i>Cisticola</i>	cyt- <i>b</i>	-	<i>Pomarea</i>	cyt- <i>b</i>	-
<i>Clytorhynchus</i>	cyt- <i>b</i>	-	<i>Prinia</i>	cyt- <i>b</i>	-
<i>Coccyzus</i>	cyt- <i>b</i>	-	<i>Progne</i>	cyt- <i>b</i>	-
<i>Colaptes</i>	cyt- <i>b</i>	-	Psittaciformes	cyt- <i>b</i>	-
Columbiformes (order)	cyt- <i>b</i>	-	<i>Pyrrhonorax</i>	cyt- <i>b</i>	-
<i>Copsychus</i>	cyt- <i>b</i>	-	<i>Pyrrhula</i>	cyt- <i>b</i>	-
<i>Corvus</i>	cyt- <i>b</i>	-	<i>Regulus</i>	cyt- <i>b</i>	-
<i>Crithagra</i> / <i>Serinus</i>	cyt- <i>b</i>	-	<i>Saxicola</i>	cyt- <i>b</i>	-
Cuculiformes (order)	cyt- <i>b</i>	-	<i>Sephanoidea</i>	cyt- <i>b</i>	-
<i>Cyanistes</i>	cyt- <i>b</i>	-	<i>Setophaga</i>	cyt- <i>b</i>	-
<i>Cyanolanius</i>	cyt- <i>b</i>	-	<i>Setophaga petechia</i>	ND2 + ATP8 + CR	109
<i>Dendrocopos</i>	cyt- <i>b</i>	-	<i>Sitta</i>	cyt- <i>b</i>	-
<i>Dicrurus</i>	cyt- <i>b</i>	-	<i>Strepera</i>	cyt- <i>b</i>	-
<i>Dumetella</i>	ND2	-	<i>Sturnus</i>	ND2	-
<i>Emberiza</i>	cyt- <i>b</i>	-	Sunbirds (family)	cyt- <i>b</i>	-
<i>Erithacus</i>	cyt- <i>b</i>	-	<i>Sylvia</i>	cyt- <i>b</i>	-
<i>Estrilda</i> / <i>Erythrura</i>	cyt- <i>b</i>	-	<i>Terpsiphone</i>	cyt- <i>b</i>	-
Finches, Galápagos Cocos	cyt- <i>b</i> + Multiple	104	<i>Troglodytes</i> / <i>Thryomanes</i>	ND2	110
<i>Foudia</i>	ATP8 + ATP6 + ND3	105	<i>Turdus</i>	cyt- <i>b</i>	-
<i>Fregilupus</i>	ND2	-	<i>Upupa</i>	cyt- <i>b</i>	-
<i>Fringilla</i>	cyt- <i>b</i>	-	<i>Vidua</i>	cyt- <i>b</i>	-
<i>Haemorrhous</i>	cyt- <i>b</i>	-	Weavers (family)	cyt- <i>b</i>	-
Hawaiian Honeycreepers	cyt- <i>b</i>	41	<i>Zosterops</i> (Indian, Atlantic)	cyt- <i>b</i> , ND3	111
<i>Horornis</i>	cyt- <i>b</i>	-			
<i>Humblotia</i>	cyt- <i>b</i>	-			
<i>Hypsipetes</i>	ND3	106			
<i>Lamprotornis</i>	ND2	-			
<i>Lanius</i>	cyt- <i>b</i>	-			
<i>Leptosomus</i>	cyt- <i>b</i>	-			
<i>Lonchura</i>	cyt- <i>b</i>	-			
<i>Loxia</i>	cyt- <i>b</i>	-			
<i>Microeca</i> / <i>Eopsaltria</i>	cyt- <i>b</i>	-			
<i>Mimus</i>	ND2	-			

Sequences were obtained from previous studies as indicated. Main source of sequences is GenBank or the new sequences produced for this study, except for the cases noted in the table, for which a matrix was directly obtained from a specific study^{41,103–111}. Details on molecular rates and molecular models applied to each alignment are provided in Supplementary Table 4.

Extended Data Table 4 | Previously published dated trees used

Taxonomic group	Source	Molecular markers	Calibration method
<i>Calyptr</i>	112	Multiple	Molecular rate
<i>Cinclodes</i>	113	3 mtDNA and 3 nuclear	Biogeographical
<i>Corvides</i>	114	Multiple	Fossils
<i>Corvus moriorum</i>	115	Mitogenome	Fossils
<i>Ducula</i>	116	ND2, COI, ND3, nuclear	Secondary & Biogeography
<i>Junco</i>	117	ND2 + CR + COI + ATP + nuclear	Molecular rate
Meliphagides (infraorder)	118	mtDNA and nuclear	Fossils & Secondary
<i>Nesillas</i>	119	ND2	Molecular rate
<i>Ptilinopus</i>	120	ND2, COI, ND3, nuclear	Secondary & Biogeography
<i>Pyrocephalus</i>	121	cyt- <i>b</i> , ND2, nuclear	Molecular rate
<i>Zosterops</i> (Pacific)	122	cyt- <i>b</i> , ND2, ND3, ATPase	Molecular rate

Data are from previously published studies¹¹²⁻¹²².

Extended Data Table 5 | Bootstrap of M14 and M19 models

Model	Cladogenesis		Extinction		Colonisation		Anagenesis	
	λ^c_0	γ	μ_0	χ	γ_0	α	λ^a_0	β
M14	0.023 (0.01 - 0.07)	0.26 (0.13 - 0.37)	1.88 (1.48 - 2.45)	0.15 (0.11 - 0.18)	51.30 (28.86 - 89.64)	0.25 (0.17 - 0.34)	0.05 (0.01 - 0.16)	0.42 (0.24 - 0.61)
M19	0.04 (0.022 - 0.077)	0.027 (0.016 - 0.034)	1.95 (1.55 - 2.50)	0.15 (0.12 - 0.18)	67.26 (36.35 - 112.71)	0.29 (0.21 - 0.37)	0.059 (0.02 - 0.19)	0.38 (0.21 - 0.57)

Maximum-likelihood estimates and 95% confidence intervals of the parameters of the two best models. Confidence intervals were obtained from the bootstrap analyses. Parameter symbols are explained in Supplementary Table 1.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Google Earth Pro v7.3.2.5776; Geneious v11; Excel v16.28.

Data analysis We added the code used for the DAISIE analyses to a new version of the DAISIE R package, which we deposited on GitHub (<https://github.com/rsetienne/DAISIE>).

Other software used: BEAST v2.4.8; jModeltest v2.1.5; R v3.5.1; RStudio v1.1.453; BBEdit v12.1.3; Cyberduck v7.1.2; Figtree v1.4.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All underlying data are available in the manuscript as supplementary data or on online databases.
New sequence data has been uploaded to GenBank with accession numbers MH307408-MH307656.

Other data types have been uploaded to Mendeley:
DNA alignments: <https://doi.org/10.17632/vf95364vx6.1>
New phylogenetic trees produced for this study: <https://doi.org/10.17632/p6hm5w8s3b.2>
DAISIE R objects: <https://doi.org/10.17632/sy58zbv3s2.2>

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We produced phylogenies for island birds and developed a new method for estimating rates of speciation, colonisation and extinction from these islands and to relate them to island area and isolation on a global scale. A total of 596 bird taxa were included in the DAISIE analyses (including those taxa for which no phylogenetic data was available but which were present on the islands). The number of individuals sampled per taxon included in the phylogenetic analyses varied between 1 and 15.
Research sample	We did not conduct experiments. Our samples are bird specimens whose DNA was used for phylogenetic analyses. Our sampling focused on native resident terrestrial birds from 41 archipelagos (listed in Figure 1) and we considered only birds that colonise by chance events. We thus excluded marine and migratory species. We focused on songbird-like and pigeon-like birds, which constitute the majority of terrestrial (land-dwelling) birds on islands. We included only species from the same trophic level: we excluded aquatic birds, birds of prey, rails and nightjars. We also excluded introduced and vagrant species. Sex and age of the individuals is not relevant for the purposes of this study. The full lists of species and samples are given in Supplementary Data 1 and Supplementary Information Table 3.
Sampling strategy	The sample size is the number of island colonisation plus island speciation events (569). We sampled all taxa of our focal group on each of 41 archipelagos.
Data collection	We sampled DNA from birds for sequencing, and compiled published sequence and phylogenetic data available. Bird samples were collected in the field by M.M., B.H.W., S.M.C., J.C.I., C.T. and L.V. New sequences were produced by K.H. and J.C.I. GenBank data and published phylogenetic trees were compiled by L.V.. New phylogenetic trees were produced by L.V.. Data on island physical features were compiled from various published sources cited in Extended Data Table 1.
Timing and spatial scale	Field work was conducted between 1999 and 2017 on the island and continental regions specified in Supplementary Data Table 3. The spatial scale is global, as field locations were located in several continents and oceans.
Data exclusions	No data were excluded.
Reproducibility	The likelihood and simulation analyses conducted in this study can be reproduced using examples provided in the R package DAISIE. We provide examples of the code and the same data used for running these analyses (e.g. see examples at the end of DAISIE_sim_global and DAISIE_MW_ML functions in the DAISIE R package).
Randomization	This is not relevant as we did not conduct experiments.
Blinding	Blinding is not relevant, as we did not conduct experiments.
Did the study involve field work?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

Field work, collection and transport

Field conditions	We conducted fieldwork on several islands worldwide in order to collect DNA samples from birds. The field conditions varied, but field work was only conducted when it was not raining to avoid hurting birds. The exact field conditions are not relevant because they do not impact the results.
Location	The 41 archipelagos/islands sampled are: Aldabra Group; Ascension; Azores; Bermuda; Canary Islands; Cape Verde; Chagos; Chatham; Christmas Island; Cocos (Costa Rica); Cocos (Keeling); Comoros; Fernando de Noronha; Galápagos; Gough; Guadalupe; Hawaii; Juan Fernández; Lord Howe; Madeira; Marianas; Marquesas; Mauritius Isl.; New Caledonia; Niue; Norfolk; Ogasawara; Palau; Pitcairn; Rapa Nui; Reunion; Rodrigues; Saint Helena; Samoa; SãoTomé e Príncipe; Selvagens; Seychelles (Inner); Society; Socorro; Tonga; Tristan da Cunha. Mainland sample locations: Angola, Andalucía (Spain), Cameroon, Equatorial Guinea, Gabon, Madagascar, Morocco. All relevant parameters (area, isolation, latitude, longitude, elevation, age) are listed in Supplementary Data 2.
Access and import/export	Information on collecting and export permits: - Angola - Biodiversity Research Protocol ISCED-Huíla and the South African National Biodiversity Institute (SANBI) (M.M.). - Cameroon - Limbe Botanical and Zoological Garden, Ministry of Scientific Research, Ministry of Forestry and Wildlife (M.M.). - Cape Verde - Cape Verde Agriculture and Environment Ministry; Ref.: 10/10 and 18/2015 (J.C.I.). - Comoros - Centre National de Documentation et de Recherche Scientifique, 2000 (B.H.W.) - Equatorial Guinea - Universidad Nacional de Guinea Ecuatorial (M.M.).

- Gabon - Centre National de la Recherche Scientifique (CENAREST), Station de Recherche de l'IRET at Ipassa-Makokou, Parc de La Lekedi, CENAREST N°AR0053/12/MENESTFPRSCJS/CG/CST/CSAR 2012 (M.M.).

- Madagascar - Ministère des Eaux et Forêts, 2002 (B.H.W.).

- Mauritius/Rodrigues - National Parks and Conservation Service (Republic of Mauritius), 1999 (B.H.W.).

- Mayotte - Direction de l'Agriculture et de la Forêt, 2000 (B.H.W.).

- Morocco - Ref: 5061/08/HCEFLCD/DLCPDN/PRN/CFF // 14-2015 (J.C.I.).

- New Caledonia, Loyalty Islands - Direction du Développement Economique, 2 Dec 2011 6101-858/PR (L.V.); 31 Jan 2014 6101-43/PR (S.M.C.).

- New Caledonia, South Province - Direction de l'Environnement Province Sud, 21 Jan 2014 Province Sud 3177-2013/ARR/DENV (S.M.C.).

- Portugal - Regional governments of:

- 1) Azores: 12/2016/DRA (J.C.I.).
- 2) Madeira: 02/2016 FAU MAD (J.C.I.).

- São Tomé e Príncipe, Direcção Geral do Ambiente, Ministério das Obras Públicas, Infraestruturas, Recursos Naturais e Ambiente 1999-present (no number) (M.M.).

- Seychelles - Bureau of Standards and Ministry of Environment, Centre National de Documentation et de Recherche Scientifique, 2000 (B.H.W.).

- Spain - Regional governments of:

- 1) Andalucía: SGYB/AF/FJRH/RE-35-36/13 (J.C.I.).
- 2) Canary Islands: Ref.: 443/02-10-2012 // Ref.: 2016/811 (J.C.I.).

- Reunion - CRBPO (Centre de Recherches sur la Biologie des Populations d'Oiseaux, Muséum National d'Histoire Naturelle, Paris), #602, 2007 (C.T and B.H.W.).

- Museum samples: Department of Ornithology and Mammalogy of the California Academy of Sciences (Laura Wilkinson & Maureen Flannery); Natural History Museum at Tring (Mark Adams); Stuttgart State Museum of Natural History.

Disturbance

Minimal disturbance to sites - we used mist-nets, which are placed temporarily and cause minimal impact.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Study did not involve laboratory animals.

Wild animals

Birds were caught in the field using mist-nets and immediately released in the same location after a blood sample was taken. No bird was injured, killed or kept captive.

The new samples collected for this study comprised 90 different species (252 individuals): *Acrocephalus rodericanus*; *Agapornis pullarius*; *Alectroenas szanzini*; *Anabathmis hartlaubii*; *Anabathmis newtonii*; *Anabathmis reichenbachii*; *Chalcophaps indica*; *Chrysococcyx cupreus*; *Chrysococcyx lucidus*; *Coccyzus melacoryphus*; *Columba larvata*; *Columba malherbii*; *Columba thomensis*; *Coracopsis vasa*; *Corvus albus*; *Crithagra burtoni*; *Crithagra capistrata*; *Crithagra mozambica*; *Crithagra rufobrunnea*; *Crithagra sulphurata*; *Cyanolanius madagascarinus*; *Cyanomitra olivacea*; *Dicrurus ludwigii*; *Dreptes thomensis*; *Erythrura psittacea*; *Erythrura trichroa*; *Estrilda astrild*; *Estrilda melpoda*; *Estrilda astrild*; *Euplectes albonotatus*; *Euplectes aureus*; *Euplectes capensis*; *Euplectes hordeaceus*; *Euplectes orix*; *Euplectes albonotatus*; *Humblotia flavirostris*; *Lanius newtoni*; *Leptosomus discolor*; *Lonchura cucullata*; *Motacilla bocagii*; *Myiagra caledonica*; *Nesoenas picturata*; *Nigrita bicolor*; *Nigrita canicapilla*; *Ploceus cucullatus*; *Ploceus grandis*; *Ploceus insignis*; *Ploceus melanogaster*; *Ploceus nigerrimus*; *Ploceus princeps*; *Ploceus sanctithomae*; *Ploceus velatus*; *Ploceus xanthops*; *Prinia malleri*; *Prinia subflava*; *Progne modesta*; *Quelea erythropus*; *Quelea quelea*; *Saxicola torquata*; *Serinus albogularis*; *Serinus citrinelloides*; *Serinus citrinipectus*; *Serinus flaviventris*; *Serinus flaviventris*; *Serinus mozambicus*; *Serinus totta*; *Streptopelia senegalensis*; *Streptopelia decaocto*; *Sylvia atricapilla*; *Sylvia borin*; *Sylvia dohrni*; *Terpsiphone atrochalybea*; *Terpsiphone rufiventer*; *Terpsiphone rufocinerea*; *Terpsiphone smithii*; *Terpsiphone viridis*; *Treron calvus*; *Treron griveaudi*; *Treron sanctithomae*; *Turdus merula*; *Turdus olivaceofuscus*; *Turdus xanthorhynchus*; *Turtur afer*; *Turtur tympanistria*; *Uraeginthus angolensis*; *Vidua macroura*; *Zosterops feae*; *Zosterops griseovirens*; *Zosterops leucophaeus*; *Zosterops lugubris*.

Sex and age of the individuals is unknown (and not relevant for this study).

Field-collected samples

Blood samples collected in the field were stored in ethanol.

Ethics oversight

No ethical approval was required as no bird was killed, injured or kept captive and we used normal procedures for mist-netting.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Stiffness of the human foot and evolution of the transverse arch

<https://doi.org/10.1038/s41586-020-2053-y>

Received: 29 March 2018

Accepted: 23 January 2020

Published online: 26 February 2020

 Check for updates

Madhusudhan Venkadesan^{1✉}, Ali Yawar¹, Carolyn M. Eng^{1,10}, Marcelo A. Dias^{2,3,4,10}, Dhiraj K. Singh^{5,9}, Steven M. Tommasini⁶, Andrew H. Haims^{6,7}, Mahesh M. Bandi^{5✉} & Shreyas Mandre^{8✉}

The stiff human foot enables an efficient push-off when walking or running, and was critical for the evolution of bipedalism^{1–6}. The uniquely arched morphology of the human midfoot is thought to stiffen it^{5–9}, whereas other primates have flat feet that bend severely in the midfoot^{7,10,11}. However, the relationship between midfoot geometry and stiffness remains debated in foot biomechanics^{12,13}, podiatry^{14,15} and palaeontology^{4–6}. These debates centre on the medial longitudinal arch^{5,6} and have not considered whether stiffness is affected by the second, transverse tarsal arch of the human foot¹⁶. Here we show that the transverse tarsal arch, acting through the inter-metatarsal tissues, is responsible for more than 40% of the longitudinal stiffness of the foot. The underlying principle resembles a floppy currency note that stiffens considerably when it curls transversally. We derive a dimensionless curvature parameter that governs the stiffness contribution of the transverse tarsal arch, demonstrate its predictive power using mechanical models of the foot and find its skeletal correlate in hominin feet. In the foot, the material properties of the inter-metatarsal tissues and the mobility of the metatarsals may additionally influence the longitudinal stiffness of the foot and thus the curvature–stiffness relationship of the transverse tarsal arch. By analysing fossils, we track the evolution of the curvature parameter among extinct hominins and show that a human-like transverse arch was a key step in the evolution of human bipedalism that predates the genus *Homo* by at least 1.5 million years. This renewed understanding of the foot may improve the clinical treatment of flatfoot disorders, the design of robotic feet and the study of foot function in locomotion.

When walking and running, people use the ball of the foot to apply forces that exceed bodyweight¹⁷. Because of these forces, the midfoot experiences large sagittal-plane torques that bend the foot. A stiff midfoot reduces the loss of propulsive work due to foot deformation and helps to efficiently utilize the mechanical power generated by the ankle during push-off^{2–4}.

The unique arch shape of the human midfoot is thought to underlie the higher stiffness of human feet compared to other primate feet^{5,6,9,18} (Extended Data Table 1). However, stiffness is not a static quantity and muscle activity can modulate midfoot stiffness in both humans and apes^{13,19,20}. The static stiffness due to the passive structures of the foot forms the baseline around which muscles with similar mechanical action as the passive tissues are likely to modulate stiffness. Therefore, understanding the morphological features underpinning the static stiffness is crucial for both static and dynamic conditions (Supplementary Information 1.1–1.3).

The human midfoot has two pronounced arches: the extensively studied medial longitudinal arch (MLA)^{5,6,20} and the less-studied

transverse tarsal arch (TTA) (Fig. 1a). The MLA stiffens the midfoot in part through a bow-string arrangement with the stiff longitudinal fibres of the plantar fascia^{7,9} and a windlass-like mechanism due to toe dorsiflexion just before push-off^{8,21}. In addition to the plantar fascia, the longitudinally oriented long plantar, short plantar and calcaneonavicular ligaments are essential for the static midfoot stiffness in humans and other primates^{9,18}. However, in contrast to the plantar fascia, the contribution of these ligaments does not depend on the height of the MLA, as shown by their nearly equal relative contributions in both arched human feet⁹ and flat monkey feet¹⁸ (Extended Data Table 1 and Supplementary Information 1.4).

The relationship between the height or curvature of the MLA and midfoot stiffness remains controversial^{5,20}. Some people have no difficulty walking with a heel-to-toe style despite having little to no MLA¹². Conflicting evidence also emerges in foot disabilities^{11,22} and surgical reconstruction of the MLA¹⁵ when correlating MLA height with foot flexibility, and casts further doubt on the relationship between the

¹Department of Mechanical Engineering and Materials Science, Yale University, New Haven, CT, USA. ²School of Science, Aalto University, Espoo, Finland. ³Nordic Institute for Theoretical Physics (NORDITA), Stockholm, Sweden. ⁴Department of Engineering, Aarhus University, Aarhus, Denmark. ⁵Nonlinear and Non-equilibrium Physics Unit, OIST Graduate University, Onna, Japan. ⁶Department of Orthopaedics and Rehabilitation, Yale University, New Haven, CT, USA. ⁷Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA.

⁸Mathematics Institute, University of Warwick, Coventry, UK. ⁹Present address: Engineering Mechanics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India.

¹⁰These authors contributed equally: Carolyn M. Eng, Marcelo A. Dias. ✉e-mail: m.venkadesan@yale.edu; bandi@oist.jp; shreyas.mandre@warwick.ac.uk

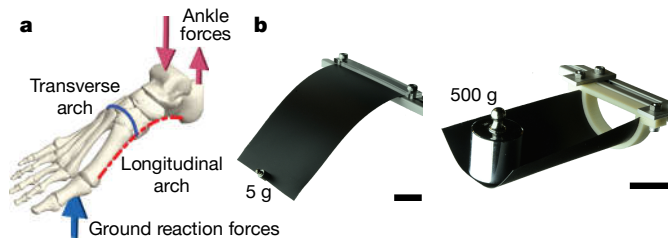


Fig. 1 | Transverse curvature and stiffness. **a**, The human foot has two distinct arches in the midfoot, the MLA and the TTA. Further anatomical details are shown in Extended Data Fig. 1. The typical loading pattern during push-off in walking and running is shown here. **b**, A thin and floppy sheet of paper becomes considerably stiffer because of transversal curvature. The TTA may have a similar role in feet. Scale bars, 5 cm.

MLA and midfoot stiffness. Furthermore, there are also debates over when a stiff midfoot arose in human evolution^{5,6}, including what kind of foot made the 3.66-million-year-old partly human-like footprints at Laetoli^{23,24}.

These debates regarding the arch morphology and stiffness centre around the MLA, the plantar fascia and other longitudinally oriented ligaments and muscles, and do not consider the role of the TTA (Supplementary Information 1.4). Even the definition of flatfoot relies mostly on the height of the MLA^{12,22}. However, the TTA may affect midfoot stiffness, similar to how even slightly curling a thin sheet of paper in the transverse direction stiffens the paper longitudinally (Fig. 1b). To investigate whether the TTA functions in this manner, we performed three-point bending tests on arched continuum shells, mechanical mimics of the midfoot and human cadaveric feet.

We investigated the relationship between curvature and stiffness by modelling the TTA as a curved elastic shell in computer simulations and physical experiments (Fig. 2a). We found that shells with greater transverse curvature were stiffer in longitudinal bending (Fig. 2b). However, the stiffness also depended on the thickness t , length L , width w , Young's modulus and Poisson's ratio of the material. To isolate the contribution of the transverse arch to midfoot stiffness, we used scaling analysis to derive dimensionless variables for stiffness and curvature that are normalized for material property and size differences (Supplementary Information 2). The normalized stiffness \hat{K} is the ratio of the stiffness of the curved shell to that of a flat plate that is identical except for the curvature. The normalized curvature \hat{c} encapsulates the mechanical coupling between bending out-of-plane and stretching in-plane that is induced by the transverse curvature c , and is given by

$$\hat{c} = \frac{cL^2}{t} \quad (1)$$

Collapse of the normalized data onto a master curve shows that \hat{c} is the chief explanatory variable for \hat{K} (Fig. 2b). There is a transition between two regimes around $\hat{c}_{tr} = 10$. Stiffness \hat{K} increases nonlinearly with curvature when $\hat{c} > \hat{c}_{tr}$ but is mostly insensitive to curvature when $\hat{c} < \hat{c}_{tr}$. Increasing the longitudinal curvature has no effect on stiffness (Fig. 2b), because these shells lack any analogue of the plantar fascia. Transverse curvature stiffens the shell because out-of-plane longitudinal bending induces in-plane stretching of the material of the shell close to the load application point (Extended Data Fig. 2 and Supplementary Information 2). Therefore, the transverse curvature has the effect of amplifying the intrinsic stiffness of a flat plate, whereas the longitudinal curvature has no similar effect.

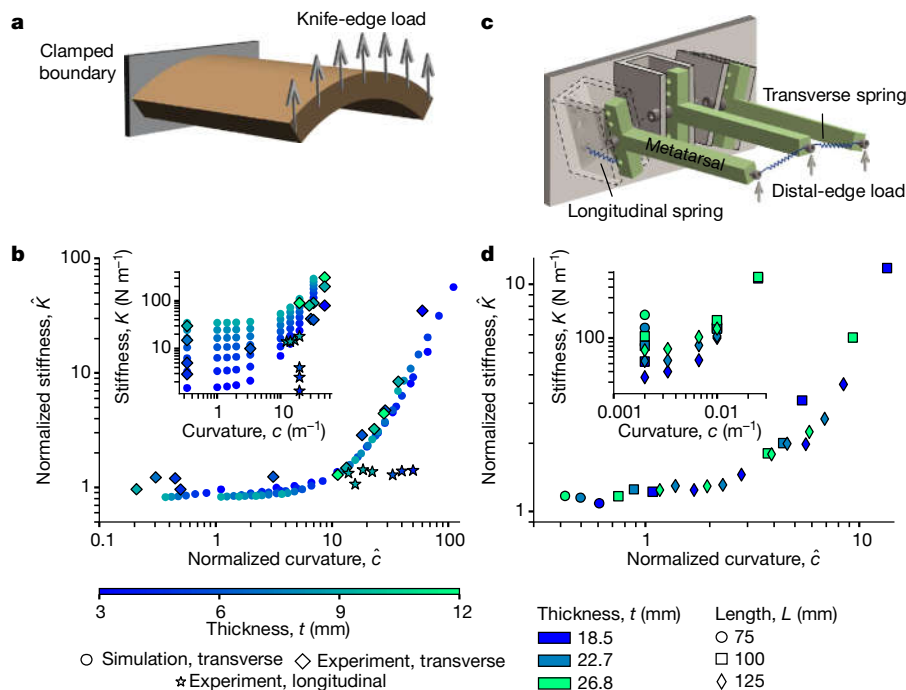


Fig. 2 | Curvature-induced stiffness in mechanical models of hominin feet.

a, Continuum elastic shells with curvature were subjected to a distributed vertical load at one end and clamped at the other. **b**, The shell data using normalized stiffness (\hat{K}) and normalized curvature (\hat{c}). The shells were transversally (diamonds) or longitudinally (stars) curved. Inset, stiffness (K) versus curvature (c) for continuum shells of various thicknesses (t) (blue shading) in experiments (diamonds and stars) and simulations (circles). **c**, The discrete foot mimics consisted of three metatarsals arranged in a transverse

arch and loaded at the distal end. Longitudinal springs at the hinged base mimic the longitudinal ligaments in feet. Transversal inter-metatarsal springs at the distal end mimic transverse elastic tissues. **d**, The foot-mimic data using normalized stiffness (\hat{K}) and normalized curvature (\hat{c}). Inset, stiffness (K) versus transverse curvature (c) for mimics of various lengths (L) and thicknesses (t). Detailed views of the continuum and discrete experiments are in Extended Data Figs. 3 and 4, respectively.

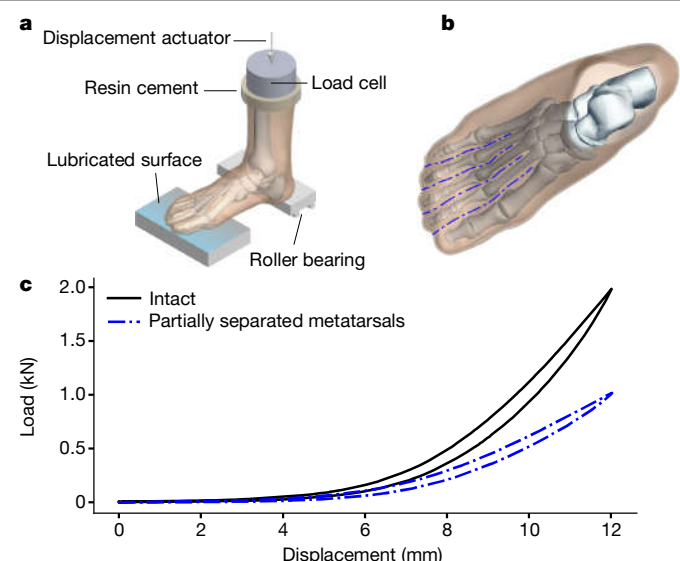


Fig. 3 | Three-point bending test on a cadaveric human foot. **a**, Fresh-frozen cadaveric feet ($n = 2$) were thawed and mounted in a materials-testing machine using an attachment at the transected shank. The distal end of the heel rested on a sliding platform with low-friction roller bearings to enable changes in foot length. The ball of the foot and the toes rested on a lubricated surface. The transected shank was displaced downward and the reaction force was measured. Tests were performed on intact feet and those with transversal cuts. **b**, The transversal cuts between the toes and metatarsals (dashed blue lines) were no deeper than the plantar plane of the metatarsal shafts. **c**, Displacement versus force traces for an intact foot (solid black line) and a foot with partially separated metatarsals (dashed blue line) foot. Some stress relaxation was observed during the initial few cycles of testing and the last cycle was used for analyses.

We performed three-point bending tests on discrete mechanical mimics of the foot with a TTA and found similar results to the continuum shells (Fig. 2c, d). The mimics, which consisted of three metatarsals

with hinges towards the midfoot, are of length L , thickness t and transverse curvature c (Methods and Supplementary Information 4). The longitudinal springs at the hinges mimic the longitudinal midfoot ligaments that contribute to midfoot stiffness whether arched or not (Supplementary Information 1.4). The distally located transverse springs mimic inter-metatarsal tissues that influence the predicted bending–stretching coupling due to the transverse curvature. We find that the normalized curvature \hat{c} accurately predicts the normalized stiffness \hat{K} for discrete foot-like structures, as for continuum shells (Fig. 2d; Methods, equation (2)). The transition in stiffness from nearly curvature-insensitive to a nonlinear increase occurs around $\hat{c}_{tr} = 3$ for the mimics. Although this value is different from continuum shells, bending–stretching coupling is the common mechanism for curvature-induced stiffness and \hat{c} emerges as the chief explanatory variable.

The role of the TTA in human feet could be found by measuring the decrease in stiffness upon flattening the TTA; however, altering the TTA would also affect other elements, such as the MLA. We therefore designed a method that emulates flattening the TTA without altering the skeletal structure. The main idea is that the transverse curvature induces stiffness by coupling longitudinal bending with stretching of the inter-metatarsal tissues, as shown by the analyses of the continuum shells and mechanical mimics, and as is also evident in mathematical models of rayed fish fins with transverse curvature²⁵. Therefore, cutting the inter-metatarsal tissues should disrupt the stiffening mechanism and emulate flattening the arch without altering the skeletal structure. We tested this idea in the foot mimics by comparing the stiffness of transversally curved mimics that lack the inter-metatarsal springs with flat mimics that had all springs intact. Both had the same stiffness ($R^2 = 0.98$, slope = 1.05, intercept = 0) (Extended Data Fig. 5), showing that cutting the transverse springs disengages the mechanism through which transverse curvature increases the longitudinal stiffness.

To determine the contribution of the TTA to stiffness in human feet, we performed three-point bending tests on two human cadaveric feet (Fig. 3a, Methods and Supplementary Information 5.2) and assessed the effect of selectively cutting the transverse tissues between the metatarsals (T condition) (Fig. 3b). To carefully preserve longitudinal tissues, we cut only the transverse metatarsal ligaments, the skin

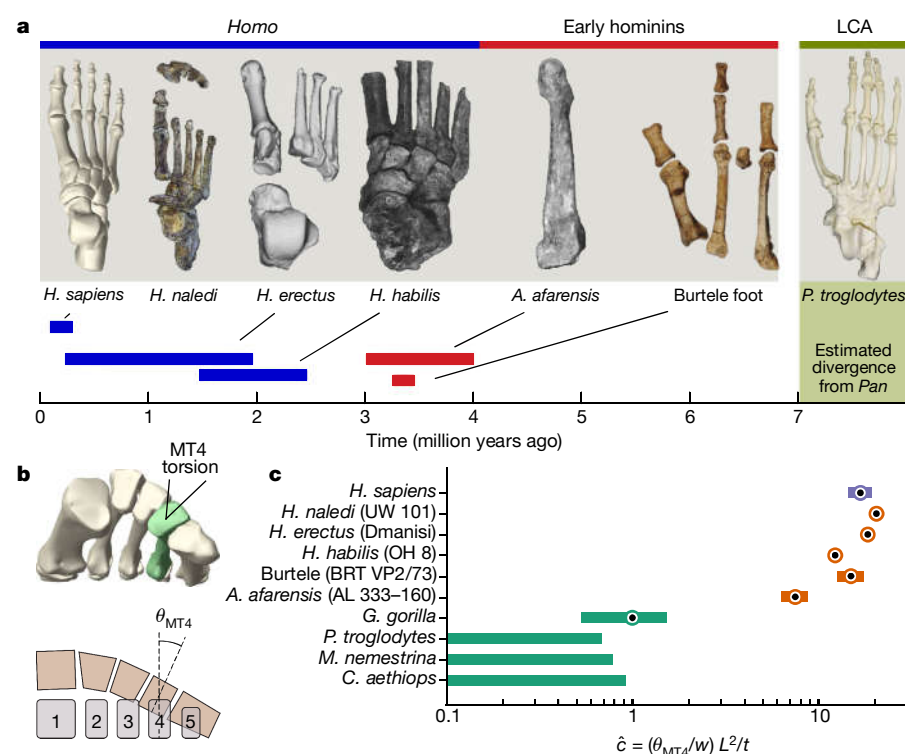


Fig. 4 | Transverse curvature of extant and extinct feet. **a**, Representative images of feet used in our analyses and their respective estimated survival dates: *H. naledi*²⁶, *H. erectus*²⁸, *H. habilis*²⁷, *A. afarensis*³⁰ and Burtele²⁹. *Pan troglodytes* represents the last common ancestor (LCA) of humans and chimpanzees. **b**, Schematics showing the skeletal view of the TTA and the torsion of the fourth metatarsal induced by the mediolateral packing of the tarso-metatarsal bones. **c**, Median (dot with circle) and the middle 50th percentile (shaded bar) of the normalized curvature (\hat{c}) are shown on a logarithmic scale. Extended Data Tables 2, 3 summarize the morphometric data used to estimate \hat{c} . Monte Carlo simulations generated the statistics for all of the samples except for the genus *Homo*. The number of independent samples used to derive the statistics: *Homo sapiens*, $n = 12$; *H. naledi*, $n = 1$; *H. erectus*, $n = 1$; *H. habilis*, $n = 1$; Burtele, $n = 1$; *A. afarensis*, $n = 1$; *Gorilla gorilla*, $n = 59$; *P. troglodytes*, $n = 106$; *Macaca nemestrina*, $n = 44$; *Chlorocebus aethiops*, $n = 56$.

between the toes and the inter-metatarsal tissues below the dorsal surface of the foot. The mechanical work to deform the foot is a measure of stiffness (Supplementary Information 5.3) and cutting these transverse tissues decreased stiffness by 44% and 54% for the two feet (Fig. 3b and Extended Data Table 1). Each foot serves as its own control, thereby quantifying the contribution of the TTA as the normalized stiffness $\hat{K} = K_{\text{intact}}/K_{\text{T-}}$. We found $\hat{K} = 1.77$ and $\hat{K} = 2.18$ for the feet for which $\hat{c} = 15.4$ and $\hat{c} = 16.0$, respectively (Fig. 4b; Methods, equation (5)).

The cadaveric experiments show that the inter-metatarsal tissues contribute substantially to foot stiffness, and more than the previously described contribution of the MLA and plantar fascia of 23% (Extended Data Table 1 and Supplementary Information 1.4). In addition to curvature of the TTA, the stiffness and slack of the inter-metatarsal tissues as well as the mobility of the metatarsals may ultimately combine to tune the longitudinal stiffness of the foot and thus influence the curvature–stiffness relationship of the TTA. Therefore, additional data are needed to find the precise curvature–stiffness relationship in human feet. Nevertheless, the mechanistic understanding of transversally curved structures suggests that the inter-metatarsal tissues affect the longitudinal bending stiffness of the foot because the human TTA, with $\hat{c} \approx 15$, is sufficiently arched to couple longitudinal bending and transverse stretching.

We use \hat{c} to compare and track the evolution of the TTA among hominins (Fig. 4 and Supplementary Information 5). At one extreme are the feet of the vervet monkey, macaque, chimpanzee and gorilla, which have $\hat{c} < 3$ and are substantially flatter than those of humans, which have $\hat{c} > 10$. At the other extreme are species in the genus *Homo*, including *Homo naledi*²⁶, *Homo habilis*²⁷ and *Homo erectus*²⁸ that possess a pronounced TTA with a human-like $\hat{c} \approx 15$. The estimated \hat{c} of the approximately 3.4-million-year-old Burtele foot (from an unidentified species) falls within the normal variation of humans despite having an abducted hallux²⁹. By contrast, the estimated \hat{c} of the approximately 3.2-million-year-old *Australopithecus afarensis* (AL-333) falls below the human range, despite a human-like torsion of the fourth metatarsal³⁰.

Additional data are needed, especially from earlier hominins such as *Ardipithecus*; however, the available evidence suggests that there were several stages in the evolution of the arch of the human foot^{5,6}. First, apes such as chimpanzees and presumably the last common ancestor of apes and hominins lack both a MLA and a TTA, and thus are able to stiffen the midfoot only partially using muscles⁵. By 3.4 million years ago, and possibly earlier, a human-like TTA had evolved that may have increased midfoot stiffness during propulsion in the Burtele hominin (Supplementary Information 5.4). Compared with humans, the TTA was apparently less developed in *A. afarensis*, which also lacked a fully developed MLA³⁰—consistent with analyses of the 3.66-million-year-old Laetoli G footprints that are thought to have been made by *A. afarensis*^{24,31}. Finally, in the genus *Homo* we see a full MLA and TTA, enabling both effective walking and running. These inferences need to be tested with additional fossils incorporating not only analyses of the MLA but also the TTA.

Our findings show a previously undescribed and substantial role for the TTA in midfoot stiffness. Traditional thinking in biomechanics, human evolution and clinical practice, with an emphasis on the sagittal plane and the MLA, should thus be expanded to incorporate the TTA and the transverse axis that is orthogonal to the sagittal plane.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2053-y>.

- Susman, R. L. Evolution of the human foot: evidence from Plio-Pleistocene hominids. *Foot Ankle* **3**, 365–376 (1983).
- Bramble, D. M. & Lieberman, D. E. Endurance running and the evolution of *Homo*. *Nature* **432**, 345–352 (2004).
- Takahashi, K. Z., Gross, M. T., van Werkhoven, H., Piazza, S. J. & Sawicki, G. S. Adding stiffness to the foot modulates soleus force-velocity behaviour during human walking. *Sci. Rep.* **6**, 29870 (2016).
- Pontzer, H. Economy and endurance in human evolution. *Curr. Biol.* **27**, R613–R621 (2017).
- Holowka, N. B. & Lieberman, D. E. Rethinking the evolution of the human foot: insights from experimental research. *J. Exp. Biol.* **221**, jeb174425 (2018).
- DeSilva, J., McNutt, E., Benoit, J. & Zipfel, B. One small step: a review of Plio-Pleistocene hominin foot evolution. *Am. J. Phys. Anthropol.* **168**, 63–140 (2019).
- Morton, D. J. Evolution of the longitudinal arch of the human foot. *J. Bone Joint Surg.* **6**, 56–90 (1924).
- Hicks, J. H. The mechanics of the foot: II. The plantar aponeurosis and the arch. *J. Anat.* **88**, 25–30 (1954).
- Ker, R. F., Bennett, M. B., Bibby, S. R., Kester, R. C. & Alexander, R. M. The spring in the arch of the human foot. *Nature* **325**, 147–149 (1987).
- D'Août, K., Aerts, P., De Clercq, D., De Meester, K. & Van Elsacker, L. Segment and joint angles of hind limb during bipedal and quadrupedal walking of the bonobo (*Pan paniscus*). *Am. J. Phys. Anthropol.* **119**, 37–51 (2002).
- Bates, K. T. et al. The evolution of compliance in the human lateral mid-foot. *Proc. R. Soc. B* **280**, 20131818 (2013).
- DeSilva, J. M. et al. Midtarsal break variation in modern humans: functional causes, skeletal correlates, and paleontological implications. *Am. J. Phys. Anthropol.* **156**, 543–552 (2015).
- Holowka, N. B., O'Neill, M. C., Thompson, N. E. & Demes, B. Chimpanzee and human midfoot motion during bipedal walking and the evolution of the longitudinal arch of the foot. *J. Hum. Evol.* **104**, 23–31 (2017).
- MacKenzie, A. J., Rome, K. & Evans, A. M. The efficacy of nonsurgical interventions for pediatric flexible flat foot: a critical review. *J. Pediatr. Orthop.* **32**, 830–834 (2012).
- Baxter, J. R. et al. Reconstruction of the medial talonavicular joint in simulated flatfoot deformity. *Foot Ankle Int.* **36**, 424–429 (2015).
- Morton, D. J. Evolution of the human foot II. *Am. J. Phys. Anthropol.* **7**, 1–52 (1924).
- Hayafune, N., Hayafune, Y. & Jacob, H. Pressure and force distribution characteristics under the normal foot during the push-off phase in gait. *Foot* **9**, 88–92 (1999).
- Bennett, M. B., Ker, R. F. & Alexander, R. M. Elastic strain energy storage in the feet of running monkeys. *J. Zool.* **217**, 469–475 (1989).
- Farris, D. J., Kelly, L. A., Cresswell, A. G. & Lichtwark, G. A. The functional importance of human foot muscles for bipedal locomotion. *Proc. Natl Acad. Sci. USA* **116**, 1645–1650 (2019).
- Heard-Booth, A. N. *Morphological and Functional Correlates of Variation in The Human Longitudinal Arch*. PhD thesis, Univ. Texas, Austin (2017).
- Griffin, N. L., Miller, C. E., Schmitt, D. & D'Août, K. Understanding the evolution of the windlass mechanism of the human foot from comparative anatomy: insights, obstacles, and future directions. *Am. J. Phys. Anthropol.* **156**, 1–10 (2015).
- Williams, D. S. & McClay, I. S. Measurements used to characterize the foot and the medial longitudinal arch: reliability and validity. *Phys. Ther.* **80**, 864–871 (2000).
- Leakey, M. D. & Hay, R. L. Pliocene footprints in the Laetoli Beds at Laetoli, northern Tanzania. *Nature* **278**, 317–323 (1979).
- Crompton, R. H. et al. Human-like external function of the foot, and fully upright gait, confirmed in the 3.66 million year old Laetoli hominin footprints by topographic statistics, experimental footprint-formation and computer simulation. *J. R. Soc. Interface* **9**, 707–719 (2012).
- Nguyen, K., Yu, N., Bandi, M. M., Venkadesan, M. & Mandre, S. Curvature-induced stiffening of a fish fin. *J. R. Soc. Interface* **14**, 20170247 (2017).
- Harcourt-Smith, W. E. H. et al. The foot of *Homo naledi*. *Nat. Commun.* **6**, 8432 (2015).
- Day, M. H. & Napier, J. R. Fossil foot bones. *Nature* **201**, 969–970 (1964).
- Pontzer, H. et al. Locomotor anatomy and biomechanics of the Dmanisi hominins. *J. Hum. Evol.* **58**, 492–504 (2010).
- Haile-Selassie, Y. et al. A new hominin foot from Ethiopia shows multiple Pliocene bipedal adaptations. *Nature* **483**, 565–569 (2012).
- Ward, C. V., Kimbel, W. H. & Johanson, D. C. Complete fourth metatarsal and arches in the foot of *Australopithecus afarensis*. *Science* **331**, 750–753 (2011).
- Raichlen, D. A., Gordon, A. D., Harcourt-Smith, W. E., Foster, A. D. & Haas, W. R. Jr. Laetoli footprints preserve earliest direct evidence of human-like bipedal biomechanics. *PLoS ONE* **5**, e9769 (2010).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ethical compliance

The authors have complied with all relevant ethical regulations in conducting the research for this paper.

Numerical simulations

We simulated the elastic response of arched shells using the Shell interface in the 3D Structural Mechanics module of COMSOL Multiphysics v.5.1 (COMSOL AB). The TTA is represented by the map for the central plane of the shell given by $\mathbf{S}_T(x, y) = (x, R_T \sin \theta_y, R_T \cos \theta_y)$ in which $\theta_y = y/R_T$, $x \in [-L/2, L/2]$ and $y \in [-w/2, w/2]$ (Extended Data Fig. 2). For all the simulations, we set $L = 0.1$ m and $w = 0.05$ m. The material was modelled as linearly elastic with Young's modulus $E = 3.5$ MPa, Poisson's ratio $\nu = 0.49$ and mass density $\rho = 965$ kg m⁻³.

The boundary at $x = -L/2$ is clamped—that is, zero displacements and rotations. The conditions at the other boundary $x = L/2$ are a uniform shear load \mathcal{T} , zero bending moment along z and zero in-plane traction so that the displacements are free (see Extended Data Fig. 2 for orientations of the axes).

We solve this model for a range of thicknesses t , from 3 mm to 9 mm in steps of 1 mm, and transverse curvature radii $R_T = 0.03$ m, 0.05 m, 0.07 m, 0.1 m, 0.3 m, 0.5 m, 0.7 m, 1 m and 3 m. For each combination of t and R_T , shear \mathcal{T} ranging from 0 N m⁻¹ to 1 N m⁻¹ is applied in increments of 0.1 N m⁻¹. The resulting out-of-plane displacement δz is measured (Extended Data Fig. 2b) and plotted against \mathcal{T} . The slope of these curves extrapolated to $\mathcal{T} = 0$ yield the stiffness defined as $k = w\mathcal{T}/\delta z$.

Continuum shell experiments

We fabricated and measured the stiffness of shells with an arch in the transverse or longitudinal directions, and compared them against a flat plate. These were all fabricated using polymer moulding techniques with polydimethylsiloxane (PDMS). The mould was fabricated using additive manufacturing (3D printed using ProJet 460Plus, 3D Systems). The printed mould was a few millimetres in thickness, with one side left open. A PDMS silicone elastomer (Sylgard 184, Dow Corning) was used to cast the arch in the mould. Because the volume ratio of the base polymer to the curing agent controls the material bulk modulus for PDMS, the same ratio of five parts base polymer to one part of curing agent by weight was consistently maintained across all fabricated arches (Supplementary Information 3). During an experiment, the fabricated arch was mounted on the experimental rig with help of clamps that were custom-fabricated to exactly match the arch curvature. The clamps were additively manufactured (Stratasys Dimension 1200es) with acrylonitrile butadiene styrene (ABSPlus) thermoplastic material (glass transition temperature, 108 °C). One end of the clamped arch was fixed to a rigid frame and the other end of the clamped arch was pushed upon by a thin edge (knife edge) that was mounted on a force sensor attached to a vertical translation stage (Extended Data Fig. 3a). The forces were measured using a data-acquisition system (LabView, National Instruments) at 2 kHz for a duration of 1 s. The load test was performed under quasi-static loading of the arch sample by providing small displacements (quasi-static steps) of 5×10^{-5} m (50 µm) per step for a total of 10 quasi-static steps (5×10^{-4} m or 500 µm). Forces were measured after each quasi-static displacement. The slope of the force–displacement curve is the stiffness K for the arch sample. Three experimental runs were conducted for each arch and their force–displacement curves were reproducible to within measurement error.

Foot mimics

We designed, fabricated and performed load–displacement tests on mechanical mimics of the foot that were transversally curved (Fig. 2, Extended Data Fig. 4 and Supplementary Information 4). The mimic consisted of three rigid metatarsals hinged at their bases. Instead of every bone in the foot, the mimics were simplifications that captured the longitudinal bending of the metatarsals and lumped all midfoot mobility into hinges at the proximal base of the metatarsals.

The metatarsals were of length L and the hinges were arranged in a transverse arch of curvature c so that the axis of each hinge was at an angle with its neighbour (Fig. 2c and Extended Data Fig. 4a). Each hinge had an extension spring held at a fixed moment arm equal to half the thickness t and provided torsional stiffness (Extended Data Fig. 4b). An inter-metatarsal transversally oriented spring connected adjacent metatarsals at the distal end and would resist any splaying induced by the transverse arch.

In hominin feet, the distal end of the metatarsals are level on the ground when loaded. Therefore, the presence of a TTA suggests increasing torsion for the lateral metatarsals (Extended Data Fig. 6b, c). The distal end of the metatarsals in the mimics were made to rest on horizontal, low-friction metallic platforms (Extended Data Fig. 4a). The vertically staggered arrangement of the platforms mimics the effect of the distal end of the metatarsals being on the same horizontal level. The platforms were attached to a micrometre-precision translation stage for applying vertical displacements. The base of the hinges were rigidly clamped to a six-axis force sensor (JR3) to measure the reaction forces due to the displacement. Stiffness was estimated as the slope of the force–displacement curve in each trial.

Multiple geometries were tested and the dimensions chosen to approximate the metatarsal lengths and midfoot widths of hominin feet, including chimpanzees and humans. The length L was varied from 75 to 125 mm (3 values), thickness t from 18.5 to 26.8 mm (3 values) and curvature from 0 to 0.025 mm⁻¹ (6 values). The spring constants, measured in an Instron materials testing machine, were 1.76 N mm⁻¹ and 0.70 N mm⁻¹ for the longitudinal and transverse springs, respectively. Three trials were performed for each foot and the force–displacement data were reproducible to within measurement error.

The normalized stiffness is $\hat{K} = K/K_{\text{flat}}$. For a flat mimic with three metatarsals, each of length L , thickness t and having a longitudinally oriented spring at its base of stiffness k_m , the longitudinal stiffness is given by $K_{\text{flat}} = 3k_m(t/2)^2/L^2$ (Supplementary Information 4.3). In a general setting, the longitudinal spring stiffness would be proportional to the width w of the midfoot by virtue of accommodating a greater amount of parallel elastic tissues. Therefore, the longitudinal stiffness is equivalently parameterized by the stiffness per unit width $k_\ell = 3k_m/w$.

Supplementary equation (4.4) for the stiffness of a flat mimic was independently verified using load–displacement tests of eight different flat mimics (Extended Data Fig. 4c and Supplementary Information 4.4). We use this relationship to normalize the measured stiffness of all of the mimics by a single chimpanzee-like flat mimic of length $L_0 = 75$ mm, thickness $t_0 = 18.5$ mm and width $w_0 = 60$ mm, and for which the measured stiffness is K_0 . By definition, the normalized stiffness of the chimpanzee-like flat mimic is $\hat{K}_0 = 1$. Therefore, the measured stiffness K of a mimic with length L , thickness t and width w is normalized according to

$$\hat{K} = \frac{K}{K_0} \left(\frac{L}{L_0} \right)^2 \left(\frac{t_0}{t} \right)^2 \left(\frac{w_0}{w} \right) \quad (2)$$

Cadaveric feet

We conducted three-point bending tests using a materials testing system (Instron model 8874) on two fresh-frozen cadaveric feet obtained from posthumous female donors (age, 55 and 64 years, body weight,

Article

1,023 N and 596 N). The loading protocol and boundary conditions under the foot were as previously described⁹. The tibia and fibula were transected midshaft and implanted in Bondo Fibreglass Resin (3M) and secured to the displacement-controlled force sensor on the Instron actuator. The ankle was at a neutral angle of 90°. The heel rested on a rigid platform that was mounted on low-friction sliders to permit foot-length changes. The forefoot rested on a highly lubricated surface to permit the foot to naturally deform in all directions when loaded. The contact point on the heel was maintained at the posterior end by placing the heel at the anterior edge of the sliding heel plate so that the heel force mimics the action of the Achilles tendon. The tests were quasi-static with a displacement rate of 0.5 mm s⁻¹ to 0.6 mm s⁻¹.

The displacement z_{peak} required to achieve a load of 3× the body weight was measured and then cyclically applied 10–15 times. The last cycle was used for analyses because there was some stress relaxation during the first 6–7 cycles. The area under the curve of the displacement z versus the force F is the work W needed to deform the foot. Following Supplementary equation (5.4), W yields an effective stiffness of the foot K_{eff} given by

$$K_{\text{eff}} = \frac{2}{z_{\text{peak}}^2} \int_0^{z_{\text{peak}}} F dz \quad (3)$$

The same measurements were repeated after bisecting the distal transverse metatarsal ligaments, the skin between the toes, and the muscles and fascia connecting the metatarsals. The inter-metatarsal tissues were transected from the dorsal surface of the foot and the cuts extended no deeper than the plantar plane of the metatarsal shafts. Therefore, none of the branches of the plantar fascia or other midfoot ligaments was affected.

Because the applied displacement was the same for the intact feet and those with bisected inter-metatarsal tissues, the ratio of work is equal to the ratio of the effective stiffness (Supplementary equation (5.5)).

Monte Carlo simulations

Anatomical variability in the size of feet (Extended Data Table 2) is incorporated using Monte Carlo simulations to generate statistics for normalized curvature (Fig. 4). The histograms generated from the Monte Carlo simulations are mostly non-Gaussian. Therefore, the median and quartiles are reported in addition to the mean and s.d. We used 1 million random combinations of the anatomical dimensions, in which each dimension was drawn from an independent Gaussian distribution with mean and s.d. values according to Extended Data Table 2, 3. Increasing the size of the Monte Carlo beyond a million samples had no effect on the statistics of the estimated quantities for the number of significant digits reported. The Monte Carlo simulations probably overestimated the variance of relevant ratios such as w/L and t/L in comparison to hominin feet, because we use independent variation of all dimensions and do not incorporate covariation that may exist. Such inflation of variance because of an assumption of independence of variables is evident when comparing primary measurements to Monte Carlo estimation of \hat{c} for humans (Extended Data Table 2).

Morphometrics of feet of extant species

Humans. Human morphometrics were obtained from 12 individuals (6 cadaveric, 6 human volunteers) using radiographic computed tomography (CT X-ray imaging) and software-based segmentation and three-dimensional model reconstruction. These feet were all evaluated by a clinical radiologist and identified as non-pathological. The collection, analyses and reporting of data from live human subjects were approved by the Yale IRB. Details on the subjects and CT data-processing methods are provided in Supplementary Information 5.1.

We measured the lever length L following the standard definition as the distance from the posterior end of the calcaneus to the anterior

end of the distal head of the third metatarsal. The width w is measured at the tarsometatarsal joint, as the mediolateral separation of the most medial aspect of the distal articular surface of the medial cuneiform to the most lateral aspect of the distal articular surface of the cuboid. The thickness t is defined as the dorso-plantar thickness of the proximal head of the third metatarsal, or the average of the second and fourth, when the third metatarsal data are unavailable. The curvature c is based on the torsion θ_{MT4} of the fourth metatarsal, which was measured using the shape of the articular surface using established protocols²⁸.

Non-human primates

Published data were used for morphometrics analysis of non-human primates: *P. troglodytes* ($n = 106$)^{28–30,32–34}, *G. gorilla* ($n = 59$)^{28–30,32–34}, *C. aethiops* ($n = 56$)^{32,35} and *M. nemestrina* ($n = 44$)^{32,36,37}.

Published data are sparse and not all required measurements were available for a single sample in the published literature for *C. aethiops* and *M. nemestrina*. Therefore, we added data from specimens that were most similar in their lever length L to the mean value reported in the literature. We carried out these measurements using software-based photogrammetry³⁸ of high-resolution images and cross-verified with measurements using a digital caliper (0.01 mm resolution). The *C. aethiops* foot is from the Yale Biological Anthropology Laboratory (YBL.3032a) and the *M. nemestrina* specimen from the Yale Peabody Museum (YPM MAM 9621).

The mean and s.d. of the lever length L were estimated from published data for chimpanzee^{32–34}, gorilla^{32,34}, *C. aethiops*^{32,35} and *M. nemestrina*^{32,36,39}. Mean w is estimated from reported w/L or dorsal skeletal views for chimpanzees and gorillas^{32,33}, and primary measurements for *C. aethiops* and *M. nemestrina*. To estimate the s.d. of w , we used reported variability in the medio-lateral width of the proximal metatarsal heads for all species^{28–30} to estimate the coefficient of variation (s.d./mean), and applied that to w . The mean and s.d. of t were all obtained from published values^{29,30} and confirmed with primary measurements for available specimens. Torsion of the fourth metatarsal θ_{MT4} is used to estimate the transverse curvature and published values were used for all non-human species included in this study^{26,28–30}. For species for which the feet are regarded as flat, we used the same metatarsal torsion values as *P. troglodytes*.

Fossil feet

We used photogrammetry³⁸ on published images of fossil feet (Fig. 4d), as well as data tables that accompanied the publication of these fossil data to estimate the necessary dimensions and ratios^{26–30}.

Among the fossil feet, all except the foot of *H. naledi*²⁶ were incomplete in some regard. For those incomplete feet, an extant species was selected as a template by taking into consideration published analyses of other postcranial and cranial elements. On the basis of this information, *H. sapiens* was chosen as the template for *H. erectus* (Dmanisi)²⁸ and *H. habilis* (Olduvai hominin)²⁷ and *G. gorilla* was chosen as the template for *A. afarensis* (AL 333)³⁰ and the unknown hominin foot found in Burtele²⁹. For example, the sole fourth metatarsal of *A. afarensis* does not permit the direct estimation of w . However, only the ratio w/L is necessary for the analyses, and the ratio of gorilla is used for the Monte Carlo analysis of the fossil. The metatarsal, however, provides a direct measurement of t , but not of L . Therefore, to estimate the ratio t/L , we incorporate the measured thickness t and the gorilla's ratio t_g/L_g by using the formula

$$\frac{t}{L} = \frac{t}{\langle t_g \rangle} \frac{t_g}{L_g} \quad (4)$$

in which $\langle t_g \rangle$ is the mean t of gorilla. This template-based estimation therefore incorporates direct measurements where available, without assuming that the fossil exactly resembles the extant template.

Curvature of hominin feet from metatarsal torsion

Following standard practice in the literature^{28,30}, we use the torsion of the fourth metatarsal (θ_{MT4}) to estimate TTA curvature. This measure also facilitates the estimation of TTA curvature using partial or disarticulated fossils. When the proximal metatarsal heads form a transverse arch and the distal metatarsal heads rest on the ground, the lateral metatarsals increasingly acquire torsion about their long axis (Fig. 4b and Extended Data Fig. 6b, c). We compared the torsion-based estimate of curvature versus using the external geometry of the dorsal surface of the skeleton and found good correspondence (Extended Data Fig. 6d and Supplementary Information 5.1). The torsion θ_{MT4} arises from the curvature c over the width w of the tarso-metatarsal articulation and therefore the curvature is approximated by $c = \theta_{\text{MT4}}/w$. Using equation (1), the torsion-based estimate of the normalized curvature parameter for the TTA is

$$\hat{c} = \frac{\theta_{\text{MT4}}}{(w/L)(t/L)} \quad (5)$$

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data supporting the findings of this study are available within the paper and its Supplementary Information.

32. Schultz, A. H. Relations between the lengths of the main parts of the foot skeleton in primates. *Folia Primatol.* **1**, 150–171 (1963).
33. Gombert, D. N. *Form and Function of The Hominoid Foot*. PhD thesis, Univ. Massachusetts, Amherst (1981).

34. Wang, W. J. & Crompton, R. H. Analysis of the human and ape foot during bipedal standing with implications for the evolution of the foot. *J. Biomech.* **37**, 1831–1836 (2004).
35. Anapol, F., Turner, T. R., Mott, C. S. & Jolly, C. J. Comparative postcranial body shape and locomotion in *Chlorocebus aethiops* and *Cercopithecus mitis*. *Am. J. Phys. Anthropol.* **127**, 231–239 (2005).
36. Sirianni, J. E., Swindler, D. R. & Tarrant, L. H. Somatometry of newborn *Macaca nemestrina*. *Folia Primatol.* **24**, 16–23 (1975).
37. Rodman, P. Skeletal differentiation of *Macaca fascicularis* and *Macaca nemestrina* in relation to arboreal and terrestrial quadrupedalism. *Am. J. Phys. Anthropol.* **51**, 51–62 (1979).
38. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
39. Hamada, Y. Standard growth patterns and variations in growth patterns of the Japanese monkeys (*Macaca fuscata*) based on an analysis by the spline function method. *Anthropol. Sci.* **102**, 57–76 (1994).
40. Hessman, F. V. Figure Calibration: a plug-in for ImageJ. http://www.astro.physik.uni-goettingen.de/~hessman/ImageJ/Figure_Calibration (2009).

Acknowledgements We thank D. Lieberman for discussions on the manuscript and S. Piazza for constructive contributions. Access to skeletal specimens was provided by G. Aronsen, K. Zyskowski, E. Sargis, Yale Biological Anthropology Laboratories and the Yale Peabody Museum. K. J. Meacham III provided experimental support. S. James helped with figures. Funding support came from the Human Frontier Science Program.

Author contributions M.V., M.M.B. and S.M. conceived the study; A.Y. and M.V. designed the foot mimics and A.Y. performed the experiments; C.M.E., A.Y. and M.V. designed, collected and analysed the data from the cadaveric experiments in consultation with S.M.T.; A.Y., C.M.E. and M.V. collected and analysed the morphometric data from cadaveric and living human feet in consultation with A.H.H.; M.A.D. and S.M. performed the mathematical modelling in consultation with M.V.; D.K.S. and M.M.B. performed the shell experiments in consultation with M.V.; M.V. planned and wrote the paper; M.V. planned and prepared the figures and tables; M.V., A.Y., S.M., M.A.D. and M.M.B. wrote the Supplementary Information; and all authors contributed to editing the paper.

Competing interests The authors declare no competing interests.

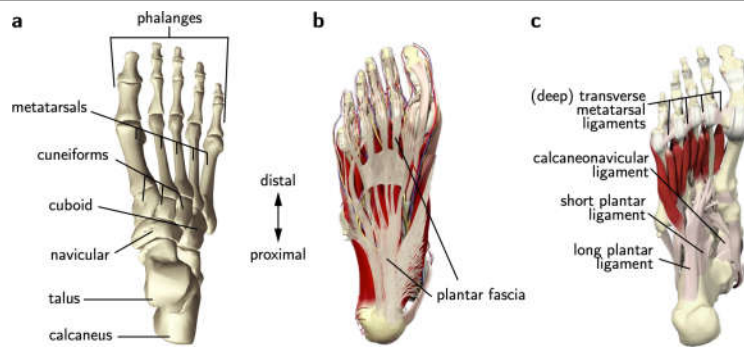
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2053-y>.

Correspondence and requests for materials should be addressed to M.V., M.M.B. or S.M.

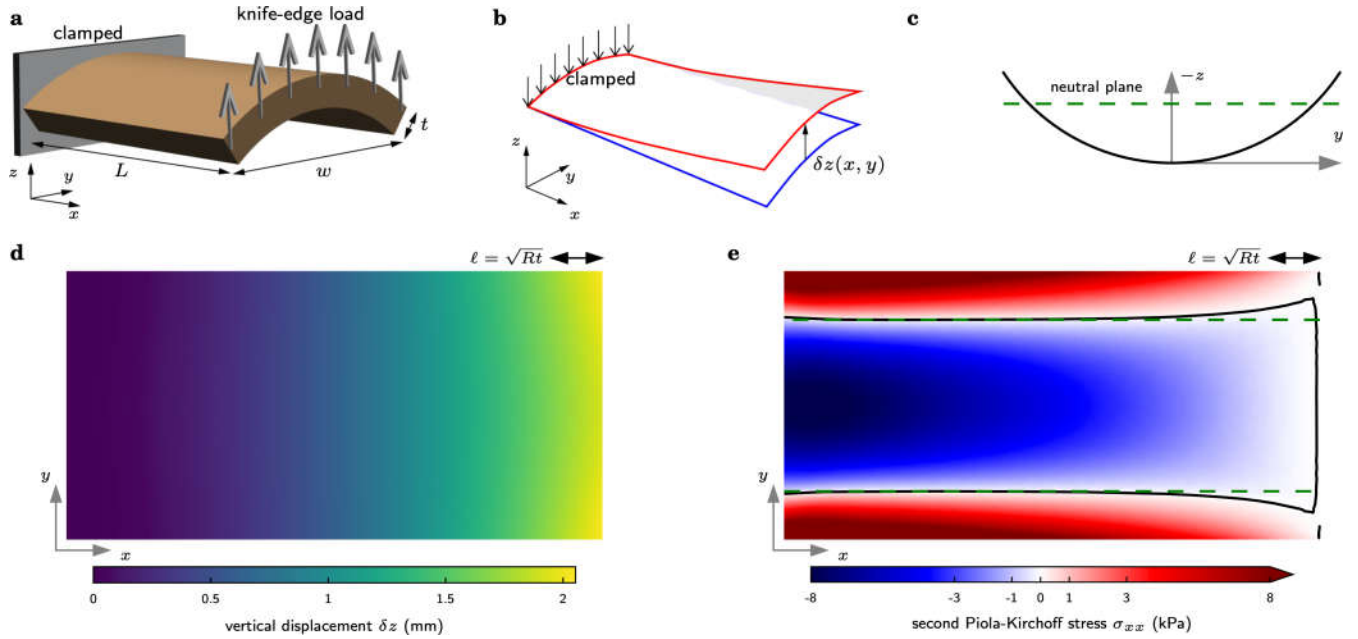
Peer review information Nature thanks Stephen Piazza and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



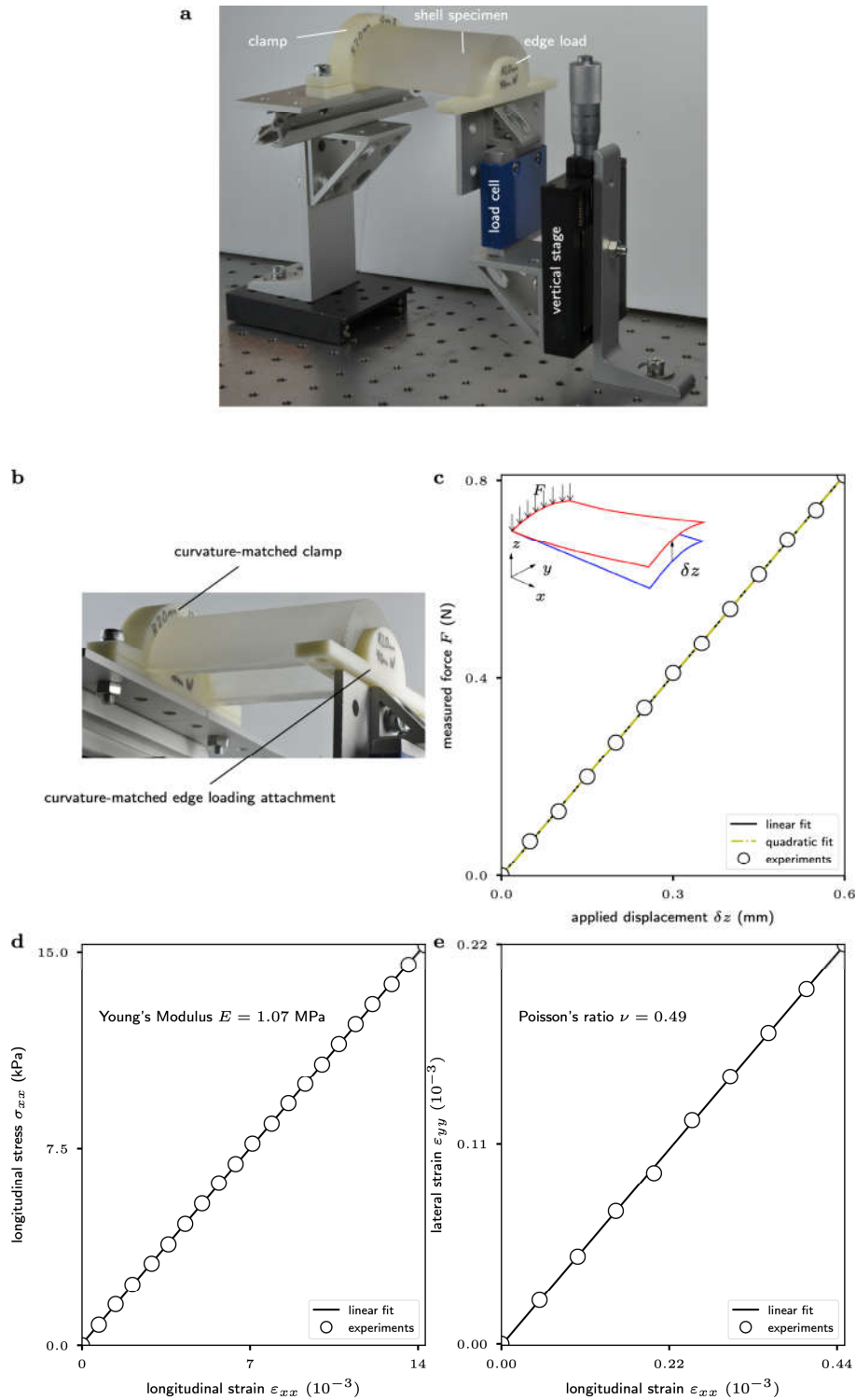
Extended Data Fig. 1 | Illustrated anatomy of the foot. **a**, Identification of the bones of the foot that are referred to in the main text. The cuneiforms, cuboid and the navicular are collectively referred to as the tarsal bones. **b**, The plantar fascia, a tough elastic band, extends from the calcaneus to the distal end of the phalanges. The fascia split and rejoin at multiple locations. **c**, The long plantar,

short plantar and calcaneonavicular ligaments are located in the midfoot and are primarily longitudinally oriented. The deep and superficial transverse metatarsal ligaments are examples of stiff, transversally oriented elastic tissues between the metatarsals. Anatomical images are from Primal Pictures.



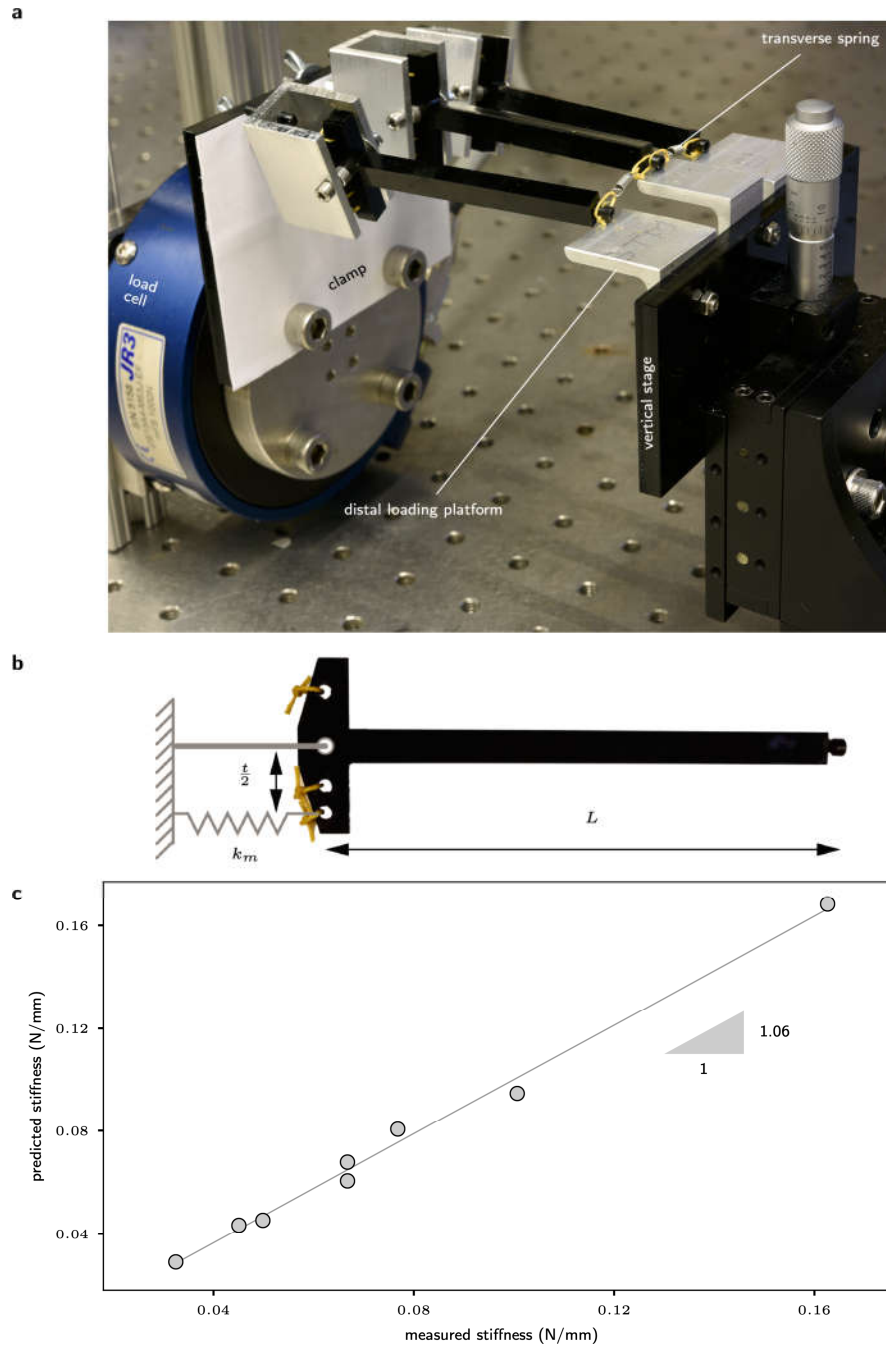
Extended Data Fig. 2 | Mathematical and computational analysis of continuum elastic shells. **a**, The shell is clamped at one end and loaded with a knife edge at the other. It is of length L , width w , thickness t and has radius of curvature R (curvature $c=1/R$). **b**, The free end displaces by a height δz on loading and reaction forces at the clamped end resist deformation. **c**, A cross-sectional view of the shell shows the location of the neutral plane, if the shell were to act as an elastic beam. **d**, Out-of-plane (z -axis) displacement profile for one numerical simulation of a shell ($L=0.1$ m, $w=0.05$ m, $t=0.003$ m,

$R=0.03$ m). Most of the displacement happens close to the loaded edge, unlike an elastic beam. **e**, The stress component σ_{xx} is shown as a colour map of the undeformed shell. In an elastic beam, the intersection of the neutral plane with the shell (**c**) would exactly match the locations of zero stress. Because of curvature-induced in-plane stretching, the zero-stress curve differs from the neutral plane predictions in the vicinity of the loaded edge and—to a lesser extent—near the clamped boundary.



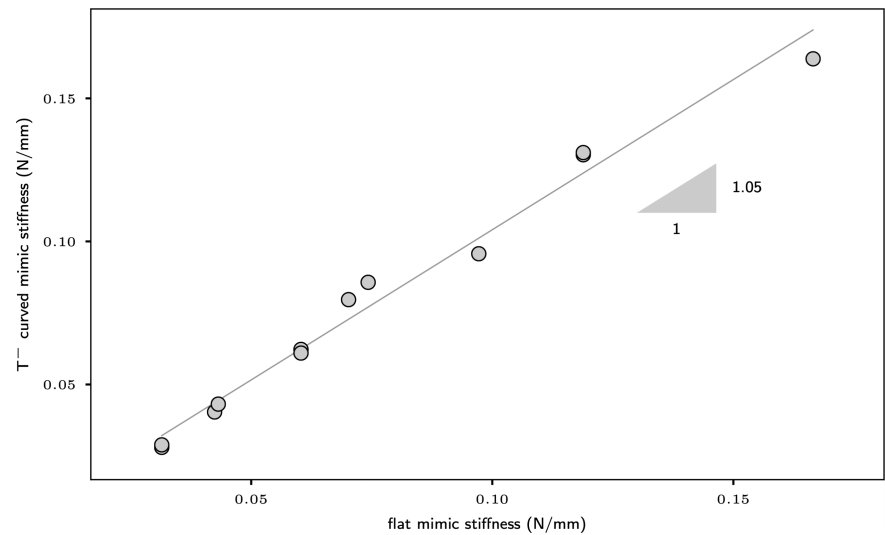
Extended Data Fig. 3 | Experimental characterization of arched shells. **a**, The experimental set-up used in stiffness measurements. **b**, A magnification of the shell from underneath shows how a curvature-matched edge-loading attachment was used to mimic a theoretical knife edge. A curvature-matched clamp was fixed and glued to the other end of the shell. **c**, Representative data

that show the linearity of the force–displacement data. The best-fit quadratic is indistinguishable from the linear fit to within sensor resolution. **d**, **e**, The Young's modulus (**d**) and Poisson's ratio (**e**) of the PDMS material used to fabricate the shells were estimated from simultaneous stress and strain measurements during an extension test of a rectangular PDMS block.

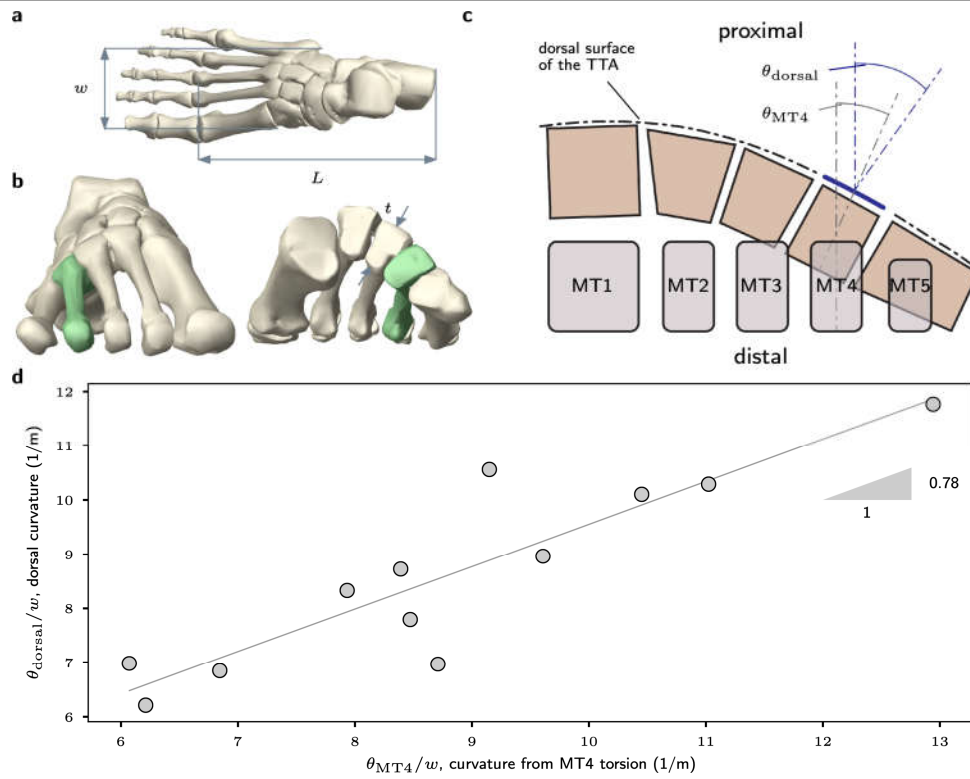


Extended Data Fig. 4 | Design and characterization of discrete mechanical foot mimics. **a**, Experimental arrangement for load–displacement measurements. The distal loading platforms for the three metatarsals are staggered in height so that all three metatarsals are loaded vertically despite the transverse curvature. In hominin feet, this is accomplished by the metatarsal torsion. **b**, Side view of a single metatarsal showing length L and thickness t of the foot mimics. The effect of thickness is to provide a moment arm for the longitudinal spring and thus affect the rotational stiffness of the

hinge. **c**, Mimics with three different thicknesses were fabricated and the thickness was estimated using load–displacement measurements on curvature-free flat mimics. The accuracy of the estimated thickness values are evaluated by plotting the predicted stiffness based on the thickness estimates against the measured stiffness. Details of the thickness estimation technique and statistics of the stiffness–stiffness correlation are provided in Supplementary Information 4.4.



Extended Data Fig. 5 | Effect of cutting the transverse springs in mechanical foot mimics. Stiffness of transversally curved foot mimics lacking the transverse inter-metatarsal springs (T-) is strongly correlated with the stiffness of flat mimics with intact transverse inter-metatarsal springs.



Extended Data Fig. 6 | Transverse curvature of hominin feet. a, Definitions of length L and width w . **b,** Definition of the thickness t . The fourth metatarsal is highlighted in green. The distal heads of the metatarsals rest flat on the ground and the proximal heads are raised away from the ground to different degrees because of the TTA. **c,** Schematic showing the accrual of torsion on the lateral metatarsals about their long axis. The curvature of the TTA was estimated using

the torsion of the fourth metatarsal θ_{MT4} . In addition, the average curvature was also estimated using the angle of the normal to the dorsal surface of the fourth metatarsal θ_{dorsal} , as measured in the midfoot (Supplementary equation (5.3)). **d,** Linear regression of the two methods to estimate TTA curvature. Details of the curvature estimation procedure and statistical results of the regression are provided in Supplementary Information 5.1.

Extended Data Table 1 | The estimated work during foot deformation

Species	Foot condition	Variable	Value
Cadaveric data collected for this study			
<i>Homo sapiens</i> , sample 1	intact	W_h	5.5 J
	–transverse tissue	W_{T-}	3.1 J
		$(W_h - W_{T-})/W_h$	44%
<i>Homo sapiens</i> , sample 2	intact	W_h	7.4 J
	–transverse tissue	W_{T-}	3.4 J
		$(W_h - W_{T-})/W_h$	54%
Previously published ^{9,18} cadaveric data			
<i>Homo sapiens</i>	intact	W_h	10.1 J
	–plantar fascia	$W_{h,pf-}$	7.8 J
		$(W_h - W_{h,pf-})/W_h$	23%
	–long plantar ligament	$W_{h,lp-}$	4.6 J
		$(W_{h,pf-} - W_{h,lp-})/W_h$	32%
	–short plantar ligament	$W_{h,sp-}$	2.1 J
		$(W_{h,lp-} - W_{h,sp-})/W_h$	25%
	–calcaneonavicular ligament	$W_{h,cn-}$	1.2 J
		$(W_{h,sp-} - W_{h,cn-})/W_h$	9%
<i>Macaca nemestrina</i>	intact	W_m	5.1 J
	plantar fascia removed	$W_{m,pf-}$	5.0 J
		$(W_m - W_{m,pf-})/W_m$	2%
	–long plantar ligament	$W_{m,lp-}$	3.1 J
		$(W_{m,pf-} - W_{m,lp-})/W_m$	37%
<i>Chlorocebus aethiops</i>	–calcaneonavicular ligament	$W_{m,cn-}$	2.4 J
		$(W_{m,lp-} - W_{m,cn-})/W_m$	14%
	intact	W_c	3.9 J

Data are obtained from cadaveric tests and from published load versus displacement data for humans⁹, *C. aethiops* and *M. nemestrina*¹⁸. In addition to the foot deformation work of the intact human foot (W_h), the cadaveric experiments performed in this study included the transection of the transverse inter-metatarsal elastic tissues, shown as W_{T-} . The peak displacements in the tests were 12 mm and 13 mm for samples 1 and 2, respectively. The published data for the three species include intact feet (W_h , W_m and W_c) and feet with transection of the plantar fascia ($W_{h,pf-}$), the long plantar ligament ($W_{h,lp-}$), the short plantar ligament ($W_{h,sp-}$) and the calcaneonavicular ligament ($W_{h,cn-}$). These estimates were obtained by digitizing the published plots^{9,18} of load versus displacement and measuring the area under the curve as the foot was loaded^{38,40}. The contribution of each of the transected tissues are represented as the ratio of the decrease in work after transection to the intact stiffness of the same foot. The previously published transections^{9,18} were performed in the same sequence as listed in this table. Raw data are available for the two cadaveric specimens as Supplementary Information.

Extended Data Table 2 | Foot morphometrics for extant species

Species	L (mm)		w (mm)		t (mm)		θ_{MT4} (deg)		\hat{c}	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
<i>Homo sapiens</i> ^a	177	16.9	50.7	4.0	16.1	1.6	25.0	4.6	16.9	2.7
<i>Homo sapiens</i> ^b	200	14.0	60.0	5.4	18.0	1.6	23.6	7.1	15.6	5.6
<i>Chlorocebus aethiops</i>	85.0	4.3	24.0	1.2	9.0	0.45	0	2.5	0.0	1.5
<i>Macaca nemestrina</i>	100	6.0	35.0	2.1	10.0	0.6	0	2.5	0.0	1.3
<i>Pan troglodytes</i>	130	13.0	52.0	5.2	13.0	1.3	0	2.5	0.0	1.2
<i>Gorilla gorilla</i>	176	17.6	72.5	7.3	16.0	1.6	2.2	1.5	1.1	0.8

^aPrimary data collected by us from 12 samples.
^bBootstrapped Monte Carlo analysis using published data.

Metrics were obtained from primary data for humans and Monte Carlo estimates for all species. For the Monte Carlo estimation, the dimensions are modelled as Gaussian random variables. Mean \pm s.d. values were obtained from values reported in the literature (see ‘Morphometrics of feet of extant species’ in the Methods for details). Although the primary data were smaller feet than the published data, the ratios w/L and t/L were almost equal. The morphometric variables are the lever length of the foot L , width of the tarso-metatarsal articular region w , dorso-plantar thickness of the third metatarsal t and torsion of the fourth metatarsal θ_{MT4} . From these, the normalized curvature parameter \hat{c} was estimated.

Extended Data Table 3 | Fossil morphometric data

Species	Specimen	θ_{MT4} (deg)	L (mm)	w (mm)	t (mm)	w/L	t/L
<i>H. naledi</i>	UW 101-1456	38.0	137.0	38.0	16.0	0.277	0.117
<i>H. erectus</i>	D2669, D4165	28.0, 29.0	—	—	17.0	$\frac{w_h}{L_h}$	$\frac{t}{\langle t_h \rangle} \frac{t_h}{L_h}$
<i>H. habilis</i>	OH 8	25.0	112	44.0	—	0.393	$\frac{t_h}{L_h}$
Burtele	BRT VP2/73	26.5	—	—	13.3	$\frac{w_g}{L_g}$	$\frac{t}{\langle t_g \rangle} \frac{t_g}{L_g}$
<i>A. afarensis</i>	AL 333-160	17.0	—	—	17.1	$\frac{w_g}{L_g}$	$\frac{t}{\langle t_g \rangle} \frac{t_g}{L_g}$

Values of L , w , t and θ_{MT4} used in estimating the normalized curvature \hat{c} of fossil samples. Variable names with a subscript h refer to human values (for example, t_h), subscript p to chimpanzee (for example, w_p) and subscript g to gorilla (for example, L_g). These values are represented by normal distributions as shown in Extended Data Table 2. Variables in angled brackets, such as $\langle t_h \rangle$, refer to the mean value shown in Extended Data Table 2. See Methods for details of source materials.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Continuum shell experiments: Labview (National Instruments)
Cadaveric experiments: FastTrack 8800 (Instron, MA, USA)
Foot mimics: Vicon Nexus 1.8 (Oxford, UK)

Data analysis

CT reconstruction and morphometrics:
* ITK-SNAP 3.6.0, Yushkevich, P. A. et al., Neuroimage 31, 1116-1128 (2006).
* 3DSlicer 4.8.1, Fedorov, A. et al., Magn Reson Imaging 30, 1323-41 (2012) www.slicer.org.
* Fiji 1.52i: Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676-82 (2012)
Elasticity simulation using COMSOL Multiphysics v5.1 (COMSOL AB, Stockholm, Sweden).
Data analyses for foot mimics: MATLAB R2019a (Mathworks, USA)
Data analyses of morphometric data: Python v3.7.5, Numpy v1.17.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study analyses previously published (and fully cited) data on foot stiffness and geometry, the details of which are part of the original publication. Other data were generated from images of unique fossil specimens or computer simulations, where sample size is not a consideration. The mechanical foot-mimics were chosen to span the relevant parameter range, which determined the sample size. Cadaveric feet were limited by availability of healthy feet and used a small sample size (n=2) similar to prevalent practice in the field. In this case, the small cadaveric sample size is justified because we directly tested a mechanism through dissections and not used the data for performing statistical estimations. Finally, we used 12 human feet for morphometrics and the sample size was determined by sample sizes that are standard practice in the field.
Data exclusions	No data were excluded.
Replication	To aid reproducibility, this study uses Monte Carlo simulations as part of a bootstrap approach to incorporate anatomical variability in the estimated quantities. Mean and standard deviations for random variables that enter the Monte Carlo are based on published (and cited) values. Furthermore, for the morphometric measurements, multiple authors independently carried out the data analyses. These efforts at replication were successful because the Monte Carlo converged, i.e. increasing number of simulations did not change the reported statistics, and the repetition of morphometric measurements by different authors yielded the consistent results.
Randomization	This study uses previously published data or unique specimen such as fossils, and therefore randomized allocation to groups is not a consideration. For the cadaveric experiments, the intact foot tests had to necessarily precede the intervention of cutting the inter-metatarsal tissues. Therefore, randomization is not possible. The measurements were automated and did not involve subjective processing and therefore blinding is not relevant.
Blinding	Blinding does not apply to this study for same reasons as randomization.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Palaeontology

Specimen provenance	All paleontological data in this study are from past publications that report on the specimen. Therefore, provenance determination is part of the past studies, which are cited in the manuscript as needed.
Specimen deposition	The data on the specimens were all from publicly available publications, and their complete citation information is provided in the manuscript.
Dating methods	No new dates are provided.
<input type="checkbox"/> Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

The human participants were comprised of 5 females and 1 male of age 35years to 68years, body height 150cm to 183cm and body mass 45kg to 97kg. The medical records and scans of these subjects were examined by a clinical radiologist and certified to be free of known diseases or deformities. The cadaveric samples were from 2 female and 4 male donors of age 55 years to 74 years, body height 160 cm to 185 cm and body mass 61 kg to 110 kg, which were amputated at the mid-tibia. The samples are obtained from consenting donors with no known foot diseases or deformities.

Recruitment

The live human data are from pre-existing CT scans of feet from patients that visited the Yale New Haven Hospital for foot diagnostics unrelated to this study. Because these subjects are from patients who had their feet scanned for clinical reasons, there is the potential for bias that they belong to a subset of the population that manifests pathologies or other changes to their feet. However, we consulted a clinical radiologist who examined a large cohort of patients and aided in the selection of only those who did not manifest any diseases or deformities. Furthermore, the clinician augmented the inspection of the radiological data with the notations on the patient's record to ensure that no cause for exclusion existed.

The cadaveric samples were from Medcure, a body donation program that is accredited by the American Association of Tissue Banks (AATB). Medcure supports medical science by providing access to body donors for researchers. In seeking foot specimens from body donors via Medcure, we used the following criteria: no foot or leg injuries or surgeries, aged between 20 and 75, no history of diabetes, and paired feet. Importantly, the specimens were not from amputees, rather from deceased individuals who died of causes unrelated to their feet. Therefore, the specimen are likely representative of healthy feet.

Ethics oversight

The Yale University IRB approved the use of CT scans collected from patients. Because the data were de-identified of all patient information, the research was granted an exemption from IRB review under federal regulation 45 CFR 46.101(b)(4). This part of the federal regulations covers research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

For the cadaveric specimens, there was no IRB approval needed since a) the subjects were not living and b) the samples were deidentified and obtained from consenting donors.

Record of consent comes from the MedCure contract (Form F-152B):

"Legal consent is obtained from all donors in accordance with the Uniform Anatomical Gift Act (UAGA) and all state and federal laws and regulations. This consent is obtained prior to specimen procurement and distribution. MEDCURE abides by Title 45, Part 46 of the U.S. Code of Federal Regulations protecting the identity of the donor."

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Neural circuitry linking mating and egg laying in *Drosophila* females

<https://doi.org/10.1038/s41586-020-2055-9>

Received: 15 May 2019

Accepted: 13 January 2020

Published online: 26 February 2020

 Check for updates

Fei Wang^{1,4}, Kaiyu Wang^{1,4}, Nora Forknall¹, Christopher Patrick¹, Tansy Yang¹, Ruchi Parekh¹, Davi Bock^{1,3} & Barry J. Dickson^{1,2}✉

Mating and egg laying are tightly coordinated events in the reproductive life of all oviparous females. Oviposition is typically rare in virgin females but is initiated after copulation. Here we identify the neural circuitry that links egg laying to mating status in *Drosophila melanogaster*. Activation of female-specific oviposition descending neurons (oviDNs) is necessary and sufficient for egg laying, and is equally potent in virgin and mated females. After mating, sex peptide—a protein from the male seminal fluid—triggers many behavioural and physiological changes in the female, including the onset of egg laying¹. Sex peptide is detected by sensory neurons in the uterus^{2–4}, and silences these neurons and their postsynaptic ascending neurons in the abdominal ganglion⁵. We show that these abdominal ganglion neurons directly activate the female-specific pC1 neurons. GABAergic (γ-aminobutyric-acid-releasing) oviposition inhibitory neurons (oviINs) mediate feed-forward inhibition from pC1 neurons to both oviDNs and their major excitatory input, the oviposition excitatory neurons (oviENs). By attenuating the abdominal ganglion inputs to pC1 neurons and oviINs, sex peptide disinhibits oviDNs to enable egg laying after mating. This circuitry thus coordinates the two key events in female reproduction: mating and egg laying.

We reasoned that egg laying is likely to depend on cell types that are female-specific and hence express one or both of the sex-determination genes⁶ *fruitless* (*fru*) and *doublesex* (*dsx*). In particular, egg laying is blocked by either silencing⁷ or masculinizing⁸ all *fru*⁺ neurons. Some of these *fru*⁺ neurons are descending interneurons, which project from the brain to the ventral nerve cord and are thought to convey high-level motor commands⁹. We therefore focused on female-specific *fru*⁺ descending neurons and used the split-GAL4 technique^{10–12} to obtain two driver lines that label two female-specific *fru*⁺*dsx*⁺ cholinergic descending neurons per brain hemisphere (Fig. 1a, b, Extended Data Figs. 1–3). In optogenetic activation experiments using Chrimson¹³, both split-GAL4 driver lines reliably induced oviposition behaviour in mated females, with most but not all females also depositing an egg (Fig. 1c, d, Supplementary Video 1; we presume that not all females had an egg in the uterus at the time of neuronal activation). Accordingly, we refer to these neurons as oviposition descending neurons (oviDNs), and to the two split-GAL4 driver lines that label them as *oviDN-SS1* and *oviDN-SS2* (in which SS denotes stable split-GAL4). Stochastic labelling of single neurons¹⁴ resolved two morphologically distinct types of oviDN, which we refer to as oviDNA and oviDNb cells (Fig. 1b). In an electron microscopy volume of a full adult female brain (FAFB¹⁵), we identified two oviDNA-like cells and one oviDNb-like cell in each hemisphere (Fig. 1b, Supplementary Video 2).

Egg laying by mated females was completely blocked by genetic ablation of oviDNs, and markedly reduced by their chronic silencing (Fig. 1e, Extended Data Fig. 4a, b). Virgin females in which oviDNs were ablated were as receptive to mating as control females (Extended Data

Fig. 4c). Several days after mating, the ovaries of oviDN-ablated females contained many mature eggs, and most carried either a fertilized egg or a first-instar larva in the uterus (Fig. 1f). We conclude that oviDNs are essential for oviposition, but dispensable for mating, ovulation and fertilization.

We were unable to generate driver lines that specifically target oviDNA or oviDNb cells. To determine which oviDN subtype is involved in oviposition, we therefore performed a stochastic ‘unsilencing’ experiment, in which a tdTomato-tagged silencing transgene was targeted to all oviDNs, but stochastically replaced in some of these cells with GFP. Individual females were assayed for egg laying over five days after mating, then dissected and stained to determine their complement of red (tdTomato; silenced) and green (GFP; unsilenced) oviDNs. Females with no unsilenced cells laid no or very few eggs, whereas those with just a single functional oviDN cell generally laid large numbers of eggs (Fig. 1g, Extended Data Fig. 5). The number of eggs laid per female was variable in these cases, but there was no appreciable difference between females in which an oviDNA cell was unsilenced and those in which an oviDNb cell was unsilenced, nor between females in which either one or two cells of either type were functional. Although the oviDNA and oviDNb subtypes differ in their morphology—and probably their connectivity and physiology—these data suggest that they nonetheless have similar functions in oviposition.

Oviposition involves a coordinated and highly stereotyped sequence of motor actions^{16,17} that progresses from abdomen bending to ovipositor extrusion and egg deposition (Fig. 2a). Abdomen bending, ovipositor extrusion and egg deposition were all eliminated in females

¹Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. ²Queensland Brain Institute, University of Queensland, St Lucia, Queensland, Australia. ³Present address: Department of Neurological Sciences, University of Vermont, Burlington, VT, USA. ⁴These authors contributed equally: Fei Wang, Kaiyu Wang. ✉e-mail: dicksonb@janelia.hhmi.org

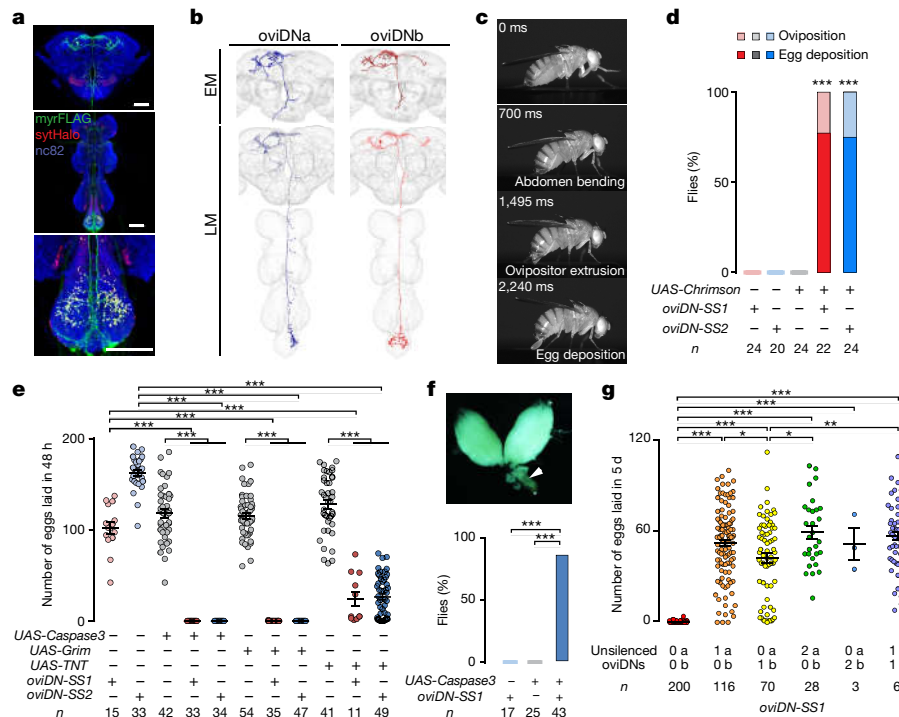


Fig. 1 | oviDNs control oviposition. **a**, Confocal images showing brain (top), ventral nerve cord (middle) and abdominal ganglion (bottom) of an *oviDN-SS1* female, stained to reveal oviDN membranes (*UAS-myrFLAG*; green), presynaptic sites (*UAS-sytHalo*; red) and all synapses (*nc82*; blue). Scale bars, 50 μ m. **b**, Electron microscopy (EM) reconstructions (top) and confocal light microscopy (LM) images (bottom) of single oviDNa and oviDNb neurons. **c**, Snapshots of the oviposition sequence that was induced after photoactivation of oviDNs (5 s, 635 nm, 261 μ W mm⁻²; Supplementary Video 1). **d**, Percentage of mated

females that exhibited oviposition and egg deposition after illumination.

e, Number of eggs laid per female in the 48 hours after mating. **f**, Percentage of females with an arrested egg (arrowhead in image) or embryo in the uterus 10 days after mating. **g**, Number of eggs laid in the five days after mating by females with all but one or two oviDNs silenced. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ by Fisher's exact test (**d**, **f**) or Wilcoxon test (**e**, **g**). Scatter plots show mean \pm s.e.m. (**e**, **g**).

in which oviDNs were ablated (Fig. 2b). Conversely, abdomen bending and ovipositor extrusion were reliably triggered by strong photoactivation of oviDNs in either virgin or mated females (Fig. 2c). Egg deposition was also induced, but only in mated females (presumably because mating is required to stimulate ovulation). In all of these oviDN activation experiments, the sequence of motor actions was the same as that in natural egg laying (Fig. 2a, c). By varying the stimulus intensity, we found that egg deposition has a higher activation threshold than abdomen bending and ovipositor extrusion (Fig. 2d), and that action latencies were shorter at higher stimulus intensities (Fig. 2e).

Moreover, at low stimulus intensities, the oviposition sequence was often truncated, but an action was never skipped, and only once did we observe a single action occurring out of order (in a total of 38 flies at each of 3 intensities; Extended Data Fig. 6). These data suggest that oviDNs may use a ramp-to-threshold mechanism to elicit the successive motor actions of oviposition¹⁸. Notably, the activation thresholds and action latencies were indistinguishable between virgins and mated females (Fig. 2d, e), indicating that mating status regulates egg laying through the brain circuits upstream of oviDNs rather than through downstream motor circuits.

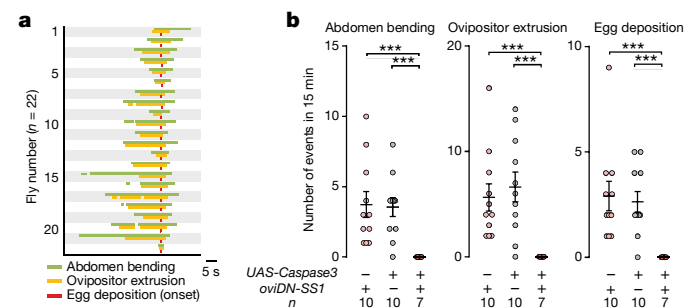


Fig. 2 | oviDNs induce the oviposition motor sequence with equal potency in virgin and mated females. **a**, Ethograms of oviposition actions by mated wild-type females, aligned to the onset of egg deposition. **b**, Frequency of oviposition motor actions by mated females. *** $P < 0.001$ by Wilcoxon test. Scatter plots show mean \pm s.e.m. **c**, Ethograms of oviposition actions that were induced by photoactivating oviDNs in mated and virgin females, aligned to the onset of the light stimulus. Pink bars indicate 2 s of 635-nm illumination.

d, **e**, Percentage of flies that exhibited the indicated actions (**d**) and latencies to action onsets, shown as mean \pm s.e.m. (**e**) at varying light intensities. Egg deposition was assessed in different flies at each light intensity ($n = 28$ –46). Other actions were examined on the same set of flies in order of increasing light intensity (virgin, $n = 14$; mated, $n = 24$). There was no significant difference between virgin and mated females (by Fisher's exact test (**d**) or Wilcoxon test (**e**)).

The onset of egg laying after mating is induced by sex peptide, a protein of the male seminal fluid¹ that is detected by sex-peptide sensory neurons (SPSNs) of the uterus^{3,4}. Sex peptide silences both SPSNs and their postsynaptic targets in the abdominal ganglion, the SAG neurons⁵. Artificially activating either SPSNs or SAG neurons suppressed egg laying in mated females^{3–5} (Fig. 3a, Extended Data Fig. 1). Conversely, ablating (Fig. 3b) or silencing^{3–5} these cells increased the number of eggs laid by virgin females. Virgin egg laying as a result of SPSN or SAG ablation depended on oviDNs, as egg laying was prevented if these cells were co-ablated (Fig. 3b). SPSN and SAG activity is thus critical in keeping oviDNs inactive until after mating. This inhibition is most likely to be indirect, because the SAGs are cholinergic and hence probably excitatory (Extended Data Fig. 3). We identified and extensively traced the ascending projections of the two SAG neurons in the FAFB volume (Fig. 3c) and found just a single synapse from SAG neurons to oviDNs (Extended Data Table 1).

The targets of SAG neurons in the brain have not been identified. Because SAG neurons regulate female receptivity as well as egg laying⁵, we speculated that their targets could include the female-specific

fru⁺dsx⁺ pC1 neurons in the protocerebrum, which are known to regulate receptivity¹⁹. Within the FAFB volume we identified five morphologically distinct pC1 cells in each hemisphere, which we refer to as pC1a–pC1e (Fig. 3d, Extended Data Fig. 7a, Supplementary Video 3). Our extensive tracing of single pC1a, pC1c and pC1e cells, as well as more limited tracing of pC1b and pC1d cells, suggests that the SAG neurons provide numerous synaptic inputs to the pC1a, pC1b and pC1c cells, with fewer if any direct inputs to pC1d and pC1e cells (Fig. 3d, Extended Data Table 1, Supplementary Video 4). We performed whole-cell recordings from individual pC1 neurons while photoactivating the SAGs (Fig. 3d–f, Extended Data Fig. 7b), and found that pC1a cells were strongly depolarized, pC1b cells were weakly depolarized and pC1c, pC1d and pC1e cells showed little or no response upon SAG activation (Fig. 3d–f). There were numerous synaptic connections amongst all five pC1 subtypes, however (Extended Data Table 1), suggesting that any information on mating status that is obtained from SAG neurons by pC1a and pC1b cells is potentially shared across the entire set of pC1 cells.

We obtained two split-GAL4 driver lines for pC1 neurons: *pC1-SS1*, which labels pC1a, pC1c and pC1e, and *pC1-SS2*, which labels all five

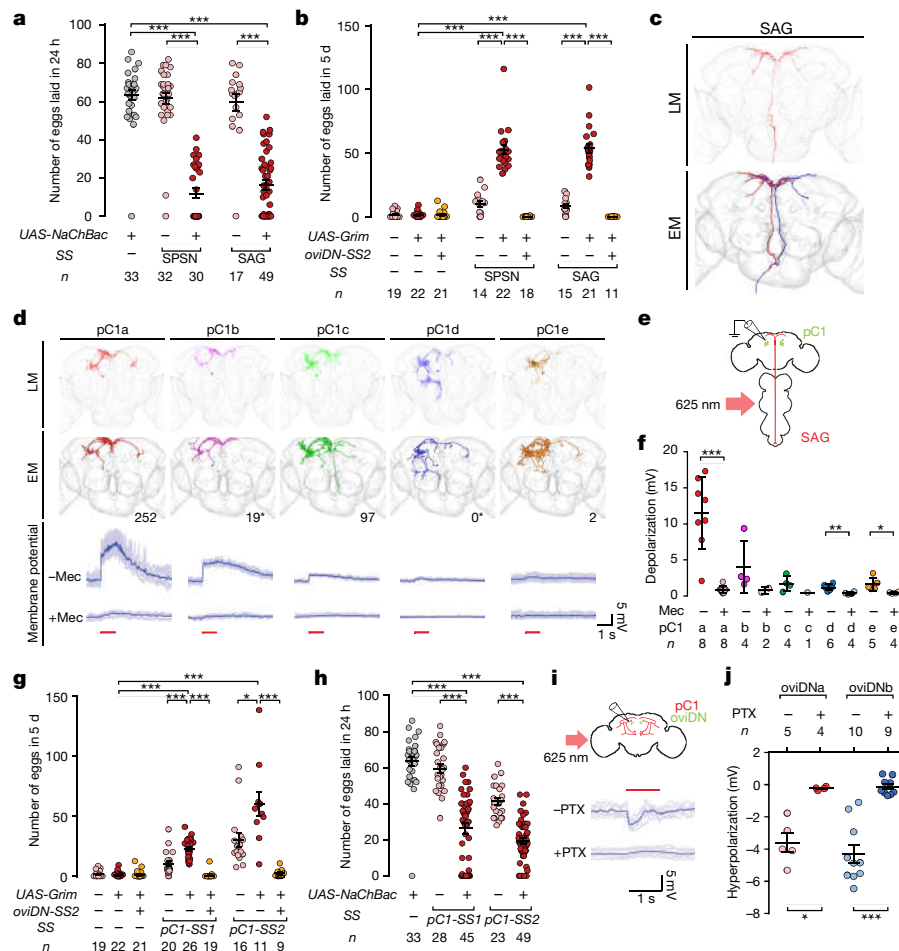


Fig. 3 | pC1 neurons suppress egg laying and oviDN activity and are regulated by the sex-peptide pathway. a, b, Number of eggs laid by mated females in the 24 hours after mating (**a**), or by virgin females in the 5 days after eclosion (**b**). **c**, Confocal image (top) and electron microscopy reconstruction (bottom) of SAG neurons. **d**, Confocal images (top) and electron microscopy reconstructions (middle) of distinct pC1 subtypes, including the number of SAG-to-pC1 synapses detected in the electron microscopy volume (asterisks indicate pC1 cells that were only partly traced). Bottom, example traces of the changes in membrane potential in pC1 cells after photoactivation of SAG neurons (1 s at 625 nm; red line) before (–Mec) and after (+Mec) mecaminamine application. Darker traces were averaged from lighter ones. **e**, Schematic of

experimental design. **f**, Peak response in each pC1 subtype (pC1a–pC1e; labelled as a–e in the figure) after photoactivation of SAG neurons, before and after mecaminamine application. **g, h**, Number of eggs laid by virgin females in the 5 days after eclosion (**g**), or by mated females in the 24 hours after mating (**h**). **i**, Schematic of experimental design (top), and example traces (bottom) showing the changes in membrane potential in an oviDNb cell after photoactivation of pC1 neurons before (–PTX) and after (+PTX) picrotoxin application. **j**, Maximum changes in oviDNa and oviDNb membrane potential in response to photoactivation of pC1 cells. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ by Wilcoxon test. Scatter plots show mean \pm s.e.m. (**a, b, f–h, j**).

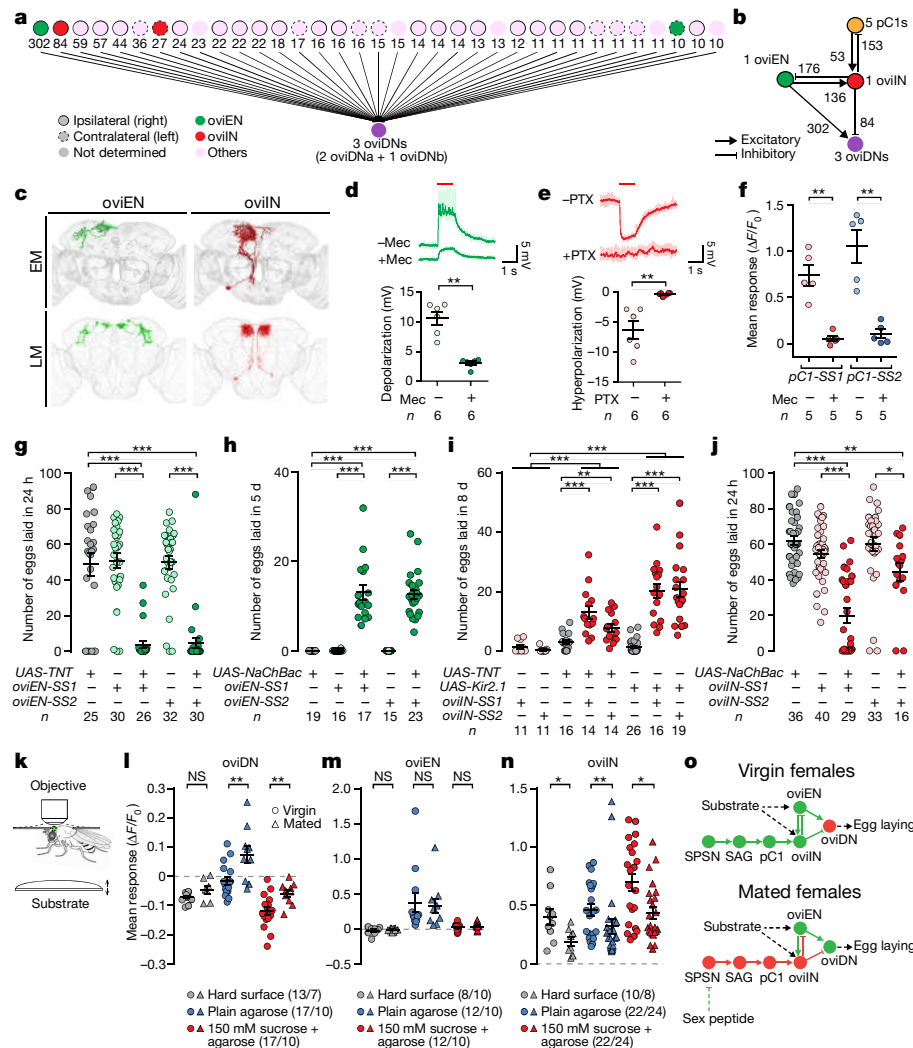


Fig. 4 | oviDNs integrate mating status and substrate signals through distinct upstream pathways. **a**, Upstream neurons of three oviDNs identified by electron microscopy reconstruction, showing the number of oviDN input synapses. **b**, Synaptic connectivity amongst four cell types in the right hemisphere. **c**, Electron microscopy reconstructions (top) and confocal images (bottom) of oviENs and oviINs. **d**, **e**, Example traces and plots of the changes in membrane potential in oviDNs that were evoked by photoactivating (1 s at 625 nm) oviENs (**d**) or oviINs (**e**), before and after application of mecamylamine or picrotoxin. **f**, Changes in the fluorescence signal of the calcium sensor GCaMP6s in oviINs in response to photoactivation of pC1 neurons. **g–j**, Number

of eggs laid by mated females in the 24 hours after mating (**g**, **j**), or by virgin females in 5 days (**h**) or 8 days (**i**) after eclosion. **k**, Schematic of the in vivo calcium-imaging experiment. The brain of a female with partially removed head cuticle (green oval) is imaged as substrates are presented to the legs sequentially using an elevator platform. **l–n**, Changes in GCaMP6s signal in oviDNs (**l**), oviENs (**m**) or oviINs (**n**) in virgin or mated females. **o**, Model for the coordination of mating and egg laying. Solid lines indicate monosynaptic connections. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ by Wilcoxon test; NS, not significant. Scatter plots show mean \pm s.e.m. (**d–j**, **l–n**).

pC1 cells (Extended Data Fig. 1). Ablation of pC1 cells using either driver resulted in an increase in egg laying in virgin females that was dependent on oviDN function (Fig. 3g), whereas mated females in which pC1 neurons were chronically activated laid fewer eggs (Fig. 3h). Brief optogenetic silencing of pC1 neurons in virgins did not acutely trigger egg laying, as would be expected if pC1-inactivated virgins (like pC1-intact mated females²⁰) rely on additional substrate-borne cues for the induction of egg laying (Extended Data Fig. 7c, d).

These behavioural data indicate that—similar to SPSNs and SAG neurons—pC1 neurons suppress the function of oviDNs and therefore suppress egg laying in virgin females. Consistent with this interpretation, we found by in vivo imaging that basal calcium levels in pC1 neurons, although variable, are generally higher in virgin than mated females (Extended Data Fig. 7e). Moreover, whole-cell recordings from oviDNs revealed that both oviDNa and oviDNb cells are hyperpolarized after photoactivation of pC1 neurons (Fig. 3i, j, Extended Data Fig. 5d),

and that this effect is sensitive to picrotoxin, a chloride channel blocker (Fig. 3i, j). This inhibition is probably indirect, because pC1 neurons are cholinergic (Extended Data Fig. 3) and have very few synapses onto the oviDNs (Extended Data Table 1).

To look for inhibitory intermediates from pC1 to oviDN cells—as well as excitatory inputs that might stimulate egg laying upon detection of a preferred substrate—we reconstructed the synaptic inputs to oviDNa and oviDNb cells in the FAFB volume (Fig. 4a, b, Extended Data Table 2). We obtained sparse split-GAL4 driver lines for the two cell types with the largest numbers of oviDN input synapses (Fig. 4c, Extended Data Fig. 1, Supplementary Video 5). Whole-cell recordings reliably showed changes in membrane potential in oviDNs after photoactivation of either of these two cell types (Fig. 4d, e). The cell type with the most oviDN input synapses is cholinergic (Extended Data Fig. 3), and activation of these cells depolarized oviDNs (Fig. 4d). We therefore named these cells oviposition excitatory neurons (oviENs). The cell type with

the second-highest number of oviDN input synapses is GABAergic (Extended Data Fig. 3), and activation of these cells hyperpolarized oviDNs (Fig. 4e). Accordingly, we named these cells oviposition inhibitory neurons (ovilNs). There is a single oviEN and a single oviLN per hemisphere, and they are reciprocally connected (Fig. 4a–c, Extended Data Table 1). The oviLNs are also reciprocally connected with pC1 cells (Fig. 4b, Extended Data Table 1), and calcium-imaging experiments showed that photoactivation of pC1 cells elicits an excitatory response in oviLNs (Fig. 4f). The pC1 cells have few direct synaptic connections with oviENs, and we did not detect any connections between SAG neurons and either oviLNs or oviENs (Extended Data Table 1).

Silencing oviENs in mated females strongly suppressed egg laying (Fig. 4g), similarly to the effect observed when oviDNs were silenced (Fig. 1e). By contrast, potentiating oviENs in virgin females caused them to lay significantly more eggs than control virgins (Fig. 4h)—albeit not as many as mated females (presumably because ovulation remains infrequent). Manipulating oviLN activity had the opposite effects: silencing oviLNs caused virgins to lay significantly more eggs (Fig. 4i), whereas depolarizing oviLNs reduced the number of eggs laid by mated females (Fig. 4j). Thus, as expected from the sign of their inputs to oviDNs (that is, excitatory for oviENs; inhibitory for oviLNs), oviENs promote egg laying, whereas oviLNs inhibit it.

We hypothesized that oviENs could mediate the external sensory signals that trigger egg laying in mated females, which are likely to include both gustatory and mechanosensory cues from the substrate²⁰. When provided with a choice of substrates, females lay more eggs on agarose medium than on a hard surface or a substrate of agarose and sucrose²¹ (Extended Data Fig. 8a–c). We therefore performed in vivo calcium imaging to determine the responses of oviDNs, oviENs and oviLNs to the presentation of each of these substrates to the legs (Fig. 4k, Extended Data Fig. 8d, e). In oviDNs, we observed an increase in calcium levels only upon contact with the agarose substrate (Fig. 4l). This response was stronger in mated females than in virgins (Fig. 4l). The agarose-and-sucrose substrate elicited a small reduction in calcium levels, which was more pronounced in virgin females (Fig. 4l). The oviENs showed a positive calcium response to agarose but to neither of the other two substrates, and this response was indistinguishable between virgins and mated females (Fig. 4m). The oviLNs responded to all three substrates, but more strongly to agarose and sucrose than to agarose alone, and only weakly to the hard surface (Fig. 4n). Regardless of substrate, oviLN responses were stronger in virgins than in mated females (Fig. 4n).

In conclusion, our findings support the following model for the neural coordination of mating and egg laying in *Drosophila* (Fig. 4o). The oviDNs control the entire oviposition motor programme. They receive excitatory input from oviENs, which respond to stimulatory cues from the substrate, and inhibitory input from oviLNs, which convey information about mating status from pC1 cells. In virgins, increased activity of pC1 neurons potentiates oviLN-mediated inhibition of both oviDNs and oviENs, which suppresses egg laying. After mating, sex peptide silences SAG inputs onto pC1 neurons, thereby decreasing the activity of pC1 neurons and oviLNs to facilitate egg laying when a preferred substrate is encountered. Reciprocal connections between oviLNs and oviENs might ensure that oviDNs respond to oviEN activation with the appropriate temporal pattern and dynamic range, through feed-forward and feedback inhibition, respectively. The oviDNs, oviENs and oviLNs all have numerous synaptic inputs in addition to those that we have described here—all of which remain functionally uncharacterized. These inputs may mediate other controls on the egg-laying process, such as the presence of an egg in the uterus²² and the nutritional state of the female²⁰. The pC1 neurons might also regulate other female behaviours that switch after mating, perhaps through different sets of output neurons. Notably, the male counterparts of pC1 neurons are thought to encode

an analogous state of courtship arousal^{23,24} that modulates command pathways for specific motor actions such as courtship song^{25,26} and ‘licking’¹⁸. Thus, functionally analogous but anatomically divergent circuits—shaped during development by *fru* and *dsx*—could account for the distinct reproductive behaviours of *Drosophila* males and females.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2055-9>.

- Kubli, E. The sex-peptide. *BioEssays* **14**, 779–784 (1992).
- Yapici, N., Kim, Y. J., Ribeiro, C. & Dickson, B. J. A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature* **451**, 33–37 (2008).
- Häsemeyer, M., Yapici, N., Heberlein, U. & Dickson, B. J. Sensory neurons in the *Drosophila* genital tract regulate female reproductive behavior. *Neuron* **61**, 511–518 (2009).
- Yang, C. H. et al. Control of the postmating behavioral switch in *Drosophila* females by internal sensory neurons. *Neuron* **61**, 519–526 (2009).
- Feng, K., Palfreyman, M. T., Häsemeyer, M., Talsma, A. & Dickson, B. J. Ascending SAG neurons control sexual receptivity of *Drosophila* females. *Neuron* **83**, 135–148 (2014).
- Auer, T. O. & Benton, R. Sexual circuitry in *Drosophila*. *Curr. Opin. Neurobiol.* **38**, 18–26 (2016).
- Kvitsiani, D. & Dickson, B. J. Shared neural circuitry for female and male sexual behaviours in *Drosophila*. *Curr. Biol.* **16**, R355–R356 (2006).
- Demir, E. & Dickson, B. J. Fruitless splicing specifies male courtship behavior in *Drosophila*. *Cell* **121**, 785–794 (2005).
- Kupfermann, I. & Weiss, K. R. The command neuron concept. *Behav. Brain Sci.* **1**, 3–10 (1978).
- Luan, H., Peabody, N. C., Vinson, C. R. & White, B. H. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron* **52**, 425–436 (2006).
- Dionne, H., Hibbard, K. L., Cavallaro, A., Kao, J. C. & Rubin, G. M. Genetic reagents for making split-GAL4 lines in *Drosophila*. *Genetics* **209**, 31–35 (2018).
- Tirian, L. & Dickson, B. J. The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the *Drosophila* nervous system. Preprint at bioRxiv <https://doi.org/10.1101/198648> (2017).
- Klapoetke, N. C. et al. Independent optical excitation of distinct neural populations. *Nat. Methods* **11**, 338–346 (2014).
- Nern, A., Pfeiffer, B. D. & Rubin, G. M. Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proc. Natl Acad. Sci. USA* **112**, E2967–E2976 (2015).
- Zheng, Z. et al. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell* **174**, 730–743 (2018).
- Yang, C. H., Belawat, P., Hafen, E., Jan, L. Y. & Jan, Y. N. *Drosophila* egg-laying site selection as a system to study simple decision-making processes. *Science* **319**, 1679–1683 (2008).
- Kimura, K., Sato, C., Koganezawa, M. & Yamamoto, D. *Drosophila* ovipositor extension in mating behavior and egg deposition involves distinct sets of brain interneurons. *PLoS ONE* **10**, e0126445 (2015).
- McKellar, C. E. et al. Threshold-based ordering of sequential actions during *Drosophila* courtship. *Curr. Biol.* **29**, 426–434 (2019).
- Zhou, C., Pan, Y., Robinett, C. C., Meissner, G. W. & Baker, B. S. Central brain neurons expressing doublesex regulate female receptivity in *Drosophila*. *Neuron* **83**, 149–163 (2014).
- Cury, K. M., Prud’homme, B. & Gompel, N. A short guide to insect oviposition: when, where and how to lay an egg. *J. Neurogenet.* **33**, 75–89 (2019).
- Gou, B., Zhu, E., He, R., Stern, U. & Yang, C. H. High throughput assay to examine egg-laying preferences of individual *Drosophila melanogaster*. *J. Vis. Exp.* **109**, e53716 (2016).
- Thomas, A. Nervous control of egg progression into the common oviduct and genital chamber of the stick-insect *Carausius morosus*. *J. Insect Physiol.* **25**, 811–823 (1979).
- Bath, D. E. et al. FlyMAD: rapid thermogenetic control of neuronal activity in freely walking *Drosophila*. *Nat. Methods* **11**, 756–762 (2014).
- Inagaki, H. K. et al. Optogenetic control of *Drosophila* using a red-shifted channelrhodopsin reveals experience-dependent influences on courtship. *Nat. Methods* **11**, 325–332 (2014).
- von Philipsborn, A. C. et al. Neuronal control of *Drosophila* courtship song. *Neuron* **69**, 509–522 (2011).
- Ribeiro, I. M. A. et al. Visual projection neurons mediating directed courtship in *Drosophila*. *Cell* **174**, 607–621 (2018).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Flies

Flies were reared on standard cornmeal–agar–molasses medium or protein-enriched food²⁷ at 25 °C with relative humidity of around 50% and a 12 h/12 h light/dark cycle, unless otherwise noted. Fly stocks used in this study are described and listed in Supplementary Tables 1 and 2.

GAL4 screen for neurons that regulate oviposition

We searched image collections of ‘generation 1’ GAL4 driver lines^{12,28} for those that potentially labelled subsets of *fru*⁺ or *dsx*⁺ neurons. Several hundred selected lines were then screened by examining light-evoked behavioural changes in females using *UAS-CsChrimson*¹³. Several GAL4 lines were found to evoke oviposition, and all labelled, amongst various other cells, a common set of descending neurons that were present in females but not males. We then sought to obtain specific split-GAL4 driver lines for these descending neurons.

Split-GAL4 screening and stabilization

Split-GAL4 lines used in this study have p65ADZp and ZpGAL4DBD inserted at the *attP40* site and *attP2* site, respectively^{11,12}, except for *pC1-SS2*, which has p65ADZp inserted at the *attP2* site and ZpGAL4DBD inserted at the first coding exon of *dsx*²⁹. p65ADZp and ZpGAL4DBD lines labelling neurons of interest were identified using a colour depth MIP (maximum intensity projection) mask search³⁰. The expression of selected combinations of p65ADZp and ZpGAL4DBD was then examined with a *UAS* reporter (*20×UAS-CsChrimson-mVenus* in *attP18*) by immunofluorescence staining and confocal microscopy (<https://www.janelia.org/project-team/flylight/protocols>). Finally, the combinations of p65ADZp and ZpGAL4DBD that gave the most specific expression patterns were stabilized by putting the two hemi-drivers in the same flies, and SS (denoting stable split-GAL4) numbers were assigned. Images of split-GAL4 lines used in this study can be viewed at <http://splitgal4.janelia.org/cgi-bin/splitgal4.cgi>.

Stochastic unsilencing

An *FRT-MCS-tdTomato-FRT* fragment was chemically synthesized (GenScript) and inserted into the *pJFRC720XUAS-IVS-mCD8::GFP* backbone. The *Kir2.1* coding sequence was synthesized (Integrated DNA Technologies) and subcloned into *20XUAS-IVS-FRT-MCS-tdTomato-FRT-mCD8::GFP* to create the in-frame *tdTomato* fusion. The resulting plasmid was inserted by phi-C31-mediated transgenesis into the *VK00005* landing sites (GenetiVision). Females carrying *hs-Flp2::PEST* in *attP3*, *UAS-FRT-Kir2.1::tdTomato-FRT-mCD8::GFP* in *VK00005* and *oviDN-SS1* or *oviDN-SS2* were heat-shocked during the first-instar larval stage or prepupal stage, with 2–4 1-hour incubations at 37 °C separated by 1-hour intervals at 22 °C. Virgin females were collected shortly after eclosion and kept in groups of 10–20 females on standard cornmeal–agar–molasses medium before being mated with wild-type males 4 days later. Females that successfully mated were then kept individually and their eggs were counted for 5 consecutive days. Females were then killed and examined for *Kir2.1::tdTomato* or *mCD8::GFP* expression in oviDNs by immunofluorescence staining and confocal microscopy.

Neuron tracing in the FAFB

Neuron skeletons in a serial section transmission electron microscopy volume of the adult female *Drosophila* brain¹⁵ were manually traced using the annotation software CATMAID³¹ (<http://www.catmaid.org>). Neuroanatomical landmarks in the electron microscopy volume such as fibre tracts, cell body size and position and neuropil boundaries

were used to search for potential candidates of the oviDNs, SAG and pC1 neurons. The process of finding relevant neurons was consistent for these cell types, relying on distinguishing features such as cell body position and tract orientation, and overall dendritic projection patterns in the confocal images. We then searched for corresponding areas of cell body position in the electron microscopy volume and followed the primary neurite emerging from the cell body as it formed fibre bundles and traversed the brain in an orientation that matched the data in the confocal images. Just enough of the primary and secondary neurites (backbone) of each potential candidate was traced to compare with confocal data, and neurons that lacked prominent morphological features in the electron microscopy volume were eliminated from consideration. Three oviDNs, five pC1 neurons and one SAG neuron were found in each hemisphere. The morphologies of oviDNs and pC1 neurons varied slightly within each group. One oviDNa, one oviDNb and the pC1a, pC1c and pC1e neurons on the right hemisphere were traced to completion. Synapses were marked on these neurons using previously described criteria for a chemical synapse¹⁵. In brief, we annotated instances in which the oviDNs, SAG and pC1 neurons were presynaptic and postsynaptic. Presynaptic locations were identified by the presence of a T-bar at an active zone with vesicles, and postsynaptic sites by the presence of postsynaptic densities (PSDs) across a synaptic cleft. At presynaptic locations in the oviDNa, SAG and pC1 neurons we identified postsynaptic neurons that contained PSDs and marked these as downstream partners; and at sites at which PSDs were present in the oviDNa, SAG and pC1 neurons, we identified locations of T-bars in the presynaptic neurons and marked these as upstream partners. Only upstream partners of oviDNb were identified and marked.

One oviIN and one oviEN were identified as upstream partners in a FAFB with the most connections to oviDNs. We traced both neurons just enough to confirm their identity. We then traced their arbors within the superior medial protocerebrum (SMP) neuropil to completion as there was extensive overlap with the oviDNs in this neuropil. Within the SMP, we identified and marked all synapses between all neurons of interest (oviIN, oviEN, three oviDNs and five pC1 neurons).

Electrophysiology

For ex vivo patch recordings, flies aged 3–5 d were immobilized on ice for around 30 s, and the nervous system was dissected out in extracellular solution³² (ECS) containing 103 mM NaCl, 3 mM KCl, 5 mM *N*-Tris(hydroxymethyl)-methyl-2-aminoethane-sulfonic acid, 10 mM trehalose, 10 mM glucose, 2 mM sucrose, 26 mM NaHCO₃, 1 mM NaH₂PO₄, 1.5 mM CaCl₂ and 4 mM MgCl₂ (pH 7.1–7.3 when bubbled with 95% (v/v) O₂/5% (v/v) CO₂, around 290 mOsm). The pia and glia sheath over the somata of interest were carefully removed with fine forceps (Domont #5SF, Fine Science Tools). The explant was subsequently mounted on a poly-D-lysine (Thermo Fisher Scientific)-coated coverslip with the somata of the target neurons facing up, and then transferred to an upright Nikon Eclipse FN1 microscope equipped with a 40×/0.8 water-immersion objective (CFI APO NIR, WD = 35 mm, Nikon). A glass micropipette with resistance of 10–15 MΩ (B150-86-7.5, Sutter Instrument) was prepared on a horizontal puller (P-1000, Sutter Instrument), and filled with intracellular solution containing 140 mM K-gluconate, 10 mM HEPES, 1 mM KCl, 4 mM MgATP, 0.5 mM Na₃GTP, 1 mM EGTA and 1% neurobiotin (SP-1120, Vector Laboratories) (pH near 7.3, around 285 mOsm). Cytoplasmic GCaMP6s³³ was expressed in the neurons of interest and visualized under 470-nm illumination to guide placement of the electrode. After obtaining a whole-cell patch, the data were collected with a Multiclamp 700B amplifier (Molecular Devices), low-pass-filtered at 2 kHz and acquired at 10 kHz with a Digidata 1440A digitizer (Molecular Devices), and analysed offline in MATLAB (MathWorks). A small hyperpolarizing current (less than 10 pA) was injected to hold the membrane potential around −65 mV. For Chrimson activation, a 625-nm-fibre-coupled LED (M625F1, Thorlabs) was placed around 5 mm away from and pointed to either the brain or the ventral

nerve cord (power intensity is about 2 mW mm^{-2}). Light stimulations were controlled by using Clampex (Molecular Devices) via the digitizer. For blocking nicotinic acetylcholine receptors or chloride channels, samples were bathed in ECS containing mecamlamine ($10 \text{ } \mu\text{M}$, M9020, Sigma-Aldrich) or picrotoxin ($150 \text{ } \mu\text{M}$, P1675, Sigma-Aldrich), respectively, for 15 min to allow the action of the antagonists.

Calcium imaging

Calcium imaging was performed at 21°C on a customized two-photon microscope equipped with a 12-kHz resonant scanner (CRS 12 KHz, Cambridge Technology), a piezo objective scanner (P-725K129, Physik Instrumente) with a controller (E-709, Physik Instrumente) and an Apo LWD $25\times/1.1$ water-immersion objective (Nikon). Z-stacks of 40 frames of either 512×512 pixels or 600×512 pixels were taken at 0.99 Hz to cover a volume of the sample, and GCaMP6s signal was captured by a photomultiplier tube (Hamamatsu Photonics) under the illumination of a two-photon laser (Chameleon, Coherent) tuned to 920 nm . The software ScanImage (Vidrio Technologies) was used to control image acquisition and synchronize stimulations. Each imaging session produced 260 volumes and lasted around 260 s. The samples were continuously perfused with ECS.

For ex vivo imaging, flies aged 4–6 days were immobilized on ice for around 30 s. The nervous system was then dissected out in ECS and mounted on a poly-D-lysine-coated coverslip. After being placed under the objective, a 625-nm -fibre-coupled LED (M625F1, Thorlabs) was placed around 5 mm away from and pointed to the brain (power intensity is around 2 mW mm^{-2}) to provide light stimulations.

For in vivo imaging, flies aged 4–6 days were immobilized on ice for about 5 min and then inserted into a rectangular hole ($1.8 \text{ mm} \times 1 \text{ mm}$) on a thin plastic sheet that was the bottom of a customized imaging chamber. The orientation of the fly's head was adjusted so that the antennae were beneath the plastic sheet, and the posterior head cuticle was facing above. Small amounts of ultraviolet (UV) curing adhesive (Loctite 352, Henkel) were applied at gaps between the fly and the hole to fix the fly in position, with a brief (around 10-s) UV irradiation (CS2010, Thorlabs). The six legs and abdomen of the fly could move freely. After filling the chamber with ECS, an observation window was opened on the head cuticle over the posterior part of the brain, and fat tissue as well as the trachea covering the posterior brain was gently removed with forceps. The oesophagus and muscles 1 and 16 were cut to minimize the movement of samples. A small plastic stage (around 1 cm in diameter) was placed at around 5 mm underneath the fly, and a manipulator (MP-285, Sutter Instrument) was used to elevate the stage to let the legs of the fly touch and stand on the stage for a period of time. The stage was covered with 1% plain agarose, 1% agarose containing 150 mM sucrose, or nothing. Each touch lasted approximately 10 s (10 volumes) with intervals of around 30 s (30 volumes).

Analysis of calcium-imaging data was done offline in Fiji³⁴ and MATLAB. In brief, z-stacks from each imaging session were averaged across all 260 time points to get a z-stack with a higher signal-to-noise ratio, which was used as a reference for identification of the neurons or neurites of interest. Then, slices covering the neurons or neurites of interest were averaged at each time point to get a time series of projection images. Sample movements during the imaging session were corrected by using TurboReg³⁵ in Fiji. Regions of interest (ROIs) were then selected by drawing polygons on the corrected time series. For each ROI, the time course of the GCaMP6s signal was obtained by averaging the fluorescence intensity of every pixel inside that ROI at each time point. The averaged fluorescence values over 10 time points before and after the onset of each stimulus were used as the baseline (F_0) and response (F), respectively. The absolute change in fluorescence intensity (ΔF) was calculated by subtracting F_0 from F , and the fluorescence intensity changes related to baseline ($\Delta F/F_0$) in each ROI were obtained.

Immunofluorescence staining

Most of the immunofluorescence staining was performed by following the standard protocols described previously³⁶. Detailed protocols for double-label staining, polarity staining and stochastic labelling in multiple colours are available at <https://www.janelia.org/project-team/flylight/protocols>. For determining the cell types that were labelled by a particular split-Gal4 driver, polarity staining was used to count the total number of cells, and stochastic labelling in multiple colours was performed to reveal the morphology of individual cells. Fluorescence in situ hybridization was performed as described previously³⁷.

For staining of Kir::tdTomato and mCD8::GFP, the central nervous system was prepared in ECS and fixed in 4% paraformaldehyde (PFA; sc-821692, Santa Cruz) at 22°C for 15 min. After being washed in phosphate-buffered saline containing 0.5% (v/v) Triton X-100 (PBT) for 30 min at 22°C , the sample was incubated in blocking buffer (S0062Z, Thermo Fisher Scientific) containing primary antibodies including rabbit anti-dsRed (1:500, 632496, Takara Bio), chicken anti-GFP (1:500, A10262, Thermo Fisher Scientific) and mouse anti-Bruchpilot (nc82, 1:25, DSHB) for 24–48 h at 4°C . The sample was then washed in PBT for 2 hours before being incubated in blocking buffer containing secondary antibodies: AF546-conjugated goat-anti-rabbit (1:300, A11035, Thermo Fisher Scientific), AF488-conjugated goat-anti-chicken (1:300, A32931, Thermo Fisher Scientific) and AF647-conjugated goat-anti-mouse (1:300, A21235, Thermo Fisher Scientific) at 4°C for 24 h. After being washed in PBT for 30 min at 22°C , the sample was dehydrated and mounted on a slide.

For staining of neurobiotin loaded into neurons during whole-cell recording, the central nervous system was dissected out and processed as described above, except that AF647-conjugated streptavidin (1:500, S21374, Thermo Fisher Scientific) was included in the primary and secondary antibodies, and AF405-conjugated goat-anti-mouse (1:300, A31553, Thermo Fisher Scientific) was used instead of AF647-conjugated goat-anti-mouse antibody.

Dehydration and DPX mounting

After incubation with secondary antibodies, the sample was washed in PBT for 15 min, fixed in 4% PFA for 10 min and sequentially dehydrated for 5 min in 30%, 50% and 75% ethanol. The sample was mounted onto a poly-D-lysine coated coverslip in 75% ethanol, and further dehydrated in 100% ethanol for 10 min. The coverslip was then submerged in xylene (X5, Thermo Fisher Scientific) for 5 min, before being mounted to a drop of DPX mountant (50-980-370, Thermo Fisher Scientific) on a slide. The slide was left to dry for 24 h before performing confocal microscopy.

Confocal microscopy and image analysis

Confocal imaging was performed under an LSM 800 or an LSM 880 inverted confocal microscope (ZEISS), with a Plan-Apochromat $20\times/0.8 \text{ M27}$ objective or a Plan-Apochromat $63\times/1.4$ oil-immersion objective (ZEISS). Images were captured using ZEN software (ZEISS), and later analysed using Fiji³⁴ and VVDViewer (https://github.com/takashi310/VVD_Viewer).

Behavioural assays and analysis

The flies used in behavioural assays were sorted and collected under light CO_2 anaesthesia 1–6 h after eclosion. Virgin females were kept in groups of 3–10 flies in vials, and males were singly housed in small food chambers ($7 \text{ mm} \times 7 \text{ mm} \times 35 \text{ mm}$). Flies used in optogenetic assays were reared on food containing 0.2 mM all-trans-retinal (Sigma-Aldrich) in darkness, before and after eclosion.

For assessing egg laying by virgin females, 4–5 flies were grouped on standard cornmeal-agar-molasses medium in single vials. The flies were transferred to new vials containing fresh food every 24 h (at around zeitgeber time (ZT) 2), and the number of eggs laid in each vial was manually counted under a stereo microscope. For assessing

Article

egg laying by mated females, virgin females were first mated with wild-type males in courtship chambers (diameter 10 mm, height 2 mm), and subsequently kept individually in vials containing protein-enriched medium, or, for the stochastic unsilencing experiments, standard cornmeal–agar–molasses medium. The number of laid eggs was counted as described above every 24 h.

For the experiment in which the position of eggs in the reproductive organs was determined, female flies were flash-frozen in liquid nitrogen and subsequently dissected in ECS under a stereo microscope. The reproductive organs were carefully uncovered by removing the cuticle over the ventral abdomen, and the presence of an egg in the uterus, oviducts and ovaries was assessed.

For examination of natural oviposition, 30 virgin females aged 4–6 d and 35 wild-type males aged 3–5 d were grouped in a food vial containing wet yeast paste, which boosts egg production while preventing the females from laying lots of eggs. After 4–5 d, single females were transferred by gentle aspiration into an observation chamber (10 mm × 30 mm × 10 mm, with a 5 mm × 10 mm groove at the centre). A small amount of cornmeal–agar–molasses medium was placed in the central groove as an egg-laying substrate. The chamber was kept in darkness with infrared illumination (880 nm) from below. The behaviour of the female around the food was videotaped from the side at a rate of 30 frames per second (fps) for 20 min.

For optogenetic activation, females were kept in darkness before being transferred into the observation chamber (diameter 10 mm or 18 mm, height 2 mm) by gentle aspiration. A customized LED panel capable of emitting infrared (880 nm) and red light (635 nm) was placed beneath the chamber to provide uniform backlight for the camera as well as red-light stimulations. The intensity and temporal pattern of light were controlled by using a customized program written in MATLAB. A camera (Manta-125C, Allied Vision) was placed above the chamber to video the behaviour of flies at 30 fps. The infrared-cut filter that came with the camera was removed to allow the detection of infrared light.

For high-speed videotaping with optogenetic activation, we modified a previously described set-up³⁸. In brief, individual females climbed upwards through a tunnel to a rectangular platform (4 mm × 2 mm) that was surrounded by a groove filled with water. The platform was illuminated by infrared light (850 nm) and was focused by LEDs providing light stimulations (5 s of continuous 625-nm illumination of 200 $\mu\text{W mm}^{-2}$). The behaviour of the fly on the platform was videoed (Ace, Basler) from the side at a rate of 200 fps.

To analyse the actions performed by female flies during natural or light-induced oviposition behaviour, videos were manually analysed offline. Three actions—abdomen bending, ovipositor extrusion and egg deposition—were analysed. Abdomen bending was defined as frames in which the abdomen was bent such that a line connecting the haltere and the abdominal tip came to meet at an angle of 15° or larger to the thoracic midline. Ovipositor extrusion was defined as any frame in which the ovipositor of the female was extruded. Egg deposition was defined as frames in which an egg was laid on the substrate.

For receptivity assays, one virgin female aged 3–6 d and one wild-type male aged 3–5 d were transferred into a courtship chamber (diameter 10 mm, height 2 mm) by gentle aspiration, and videoed under white-light illumination for a period of 30 min. The copulation rate was checked every 2 min.

For egg-laying preference assays, we adapted a set-up developed previously²¹. In brief, 30 virgin females aged 4–6 d and 35 wild-type males aged 3–5 d were grouped in a food vial containing wet yeast paste, which boosts egg production but limits egg deposition. After 4–5 d,

single females were gently aspirated into the observation chamber, which contains two 1% agarose grooves, one with and one without sucrose. The female's behaviour was videoed for 12 h and the eggs laid were counted. The preference index was calculated as the difference between egg numbers on two grooves divided by the total number of eggs.

Statistics

All egg-laying and electrophysiology data were analysed by unpaired Wilcoxon signed-rank test. Egg-position data were analysed by Fisher's exact test. All of the statistical analyses were performed using R software or MATLAB.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

- Backhaus, B., Sulkowski, E. & Schlote, F. W. A semi-synthetic, general-purpose medium for *Drosophila melanogaster*. *Drosoph. Inf. Serv.* **60**, 210–212 (1984).
- Jenett, A. et al. A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991–1001 (2012).
- Shirangi, T. R., Wong, A. M., Truman, J. W. & Stern, D. L. *Doublesex* regulates the connectivity of a neural circuit controlling *Drosophila* male courtship song. *Dev. Cell* **37**, 533–544 (2016).
- Otsuna, H., Ito, M. & Kawase, T. Color depth MIP mask search: a new tool to expedite Split-GAL4 creation. Preprint at bioRxiv <https://doi.org/10.1101/318006> (2018).
- Schneider-Mizell, C. M. et al. Quantitative neuroanatomy for connectomics in *Drosophila*. *eLife* **5**, e12059 (2016).
- Wilson, R. I. & Laurent, G. Role of GABAergic inhibition in shaping odor-evoked spatiotemporal patterns in the *Drosophila* antennal lobe. *J. Neurosci.* **25**, 9069–9079 (2005).
- Chen, T. W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
- Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
- Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
- Wu, M. et al. Visual projection neurons in the *Drosophila* lobula link feature detection to distinct behavioral programs. *eLife* **5**, e21022 (2016).
- Meissner, G. W. et al. Mapping neurotransmitter identity in the whole-mount *Drosophila* brain using multiplex high-throughput fluorescence *in situ* hybridization. *Genetics* **211**, 473–482 (2019).
- von Reyn, C. R. et al. A spike-timing mechanism for action selection. *Nat. Neurosci.* **17**, 962–970 (2014).

Acknowledgements We thank the Janelia FlyLight, Fly Facility, Project Technical Resources, Molecular Biology, Functional Connectome and Experimental Technology teams for technical assistance; A. Edmonson-Stait and G. Jefferis for initial tracing of one of the oviIN cells; U. Heberlein, K. Feng and V. Vijayan for comments on the manuscript; and V. Vijayan and G. Maimon for sharing preliminary oviDN calcium-imaging data. This work was funded by the Howard Hughes Medical Institute.

Author contributions B.J.D., F.W. and K.W. conceived the study and wrote the manuscript. F.W. and K.W. performed all experiments and analysed the data. N.F., C.P., T.Y., F.W. and R.P. reconstructed selected neurons and synapses in the FAFB electron microscopy volume, which was provided before publication by D.B.

Competing interests The authors declare no competing interests.

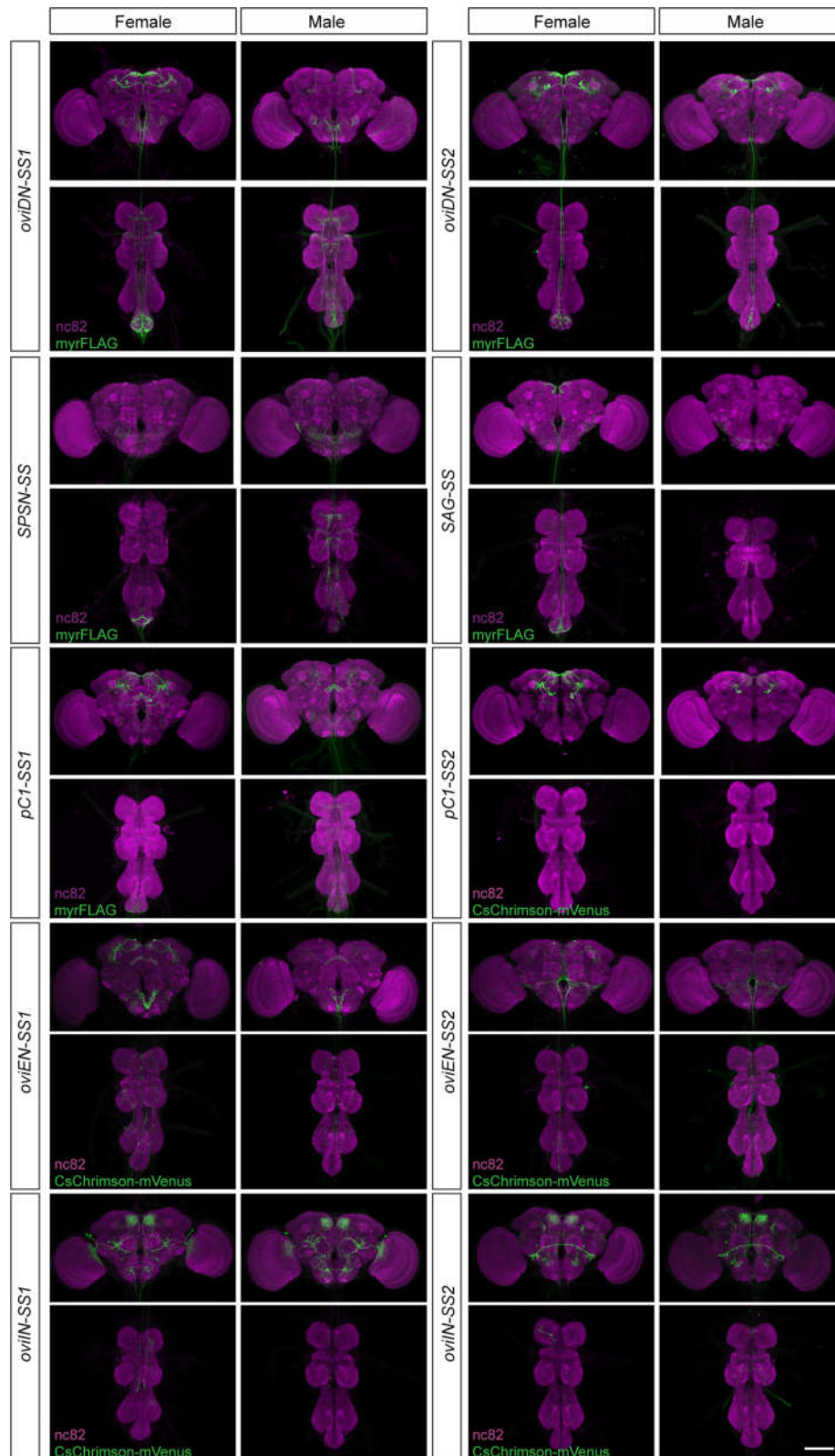
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2055-9>.

Correspondence and requests for materials should be addressed to B.J.D.

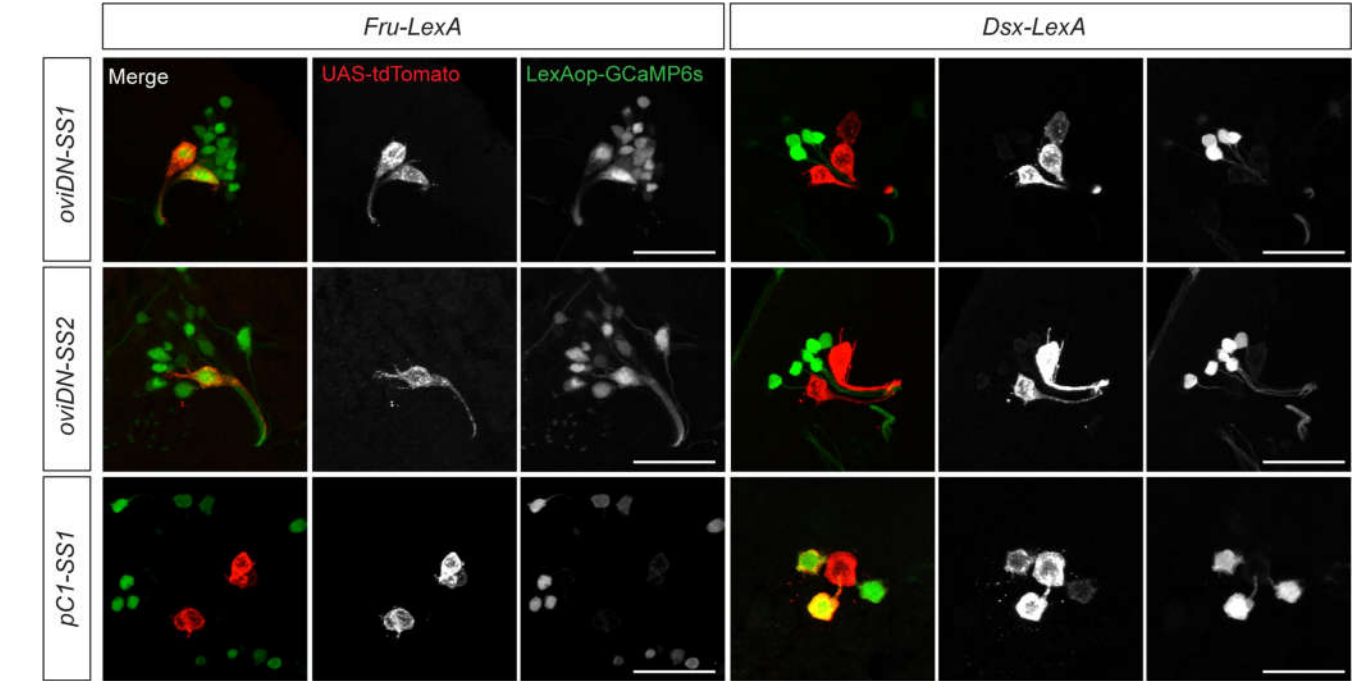
Peer review information Nature thanks Rebecca Yang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

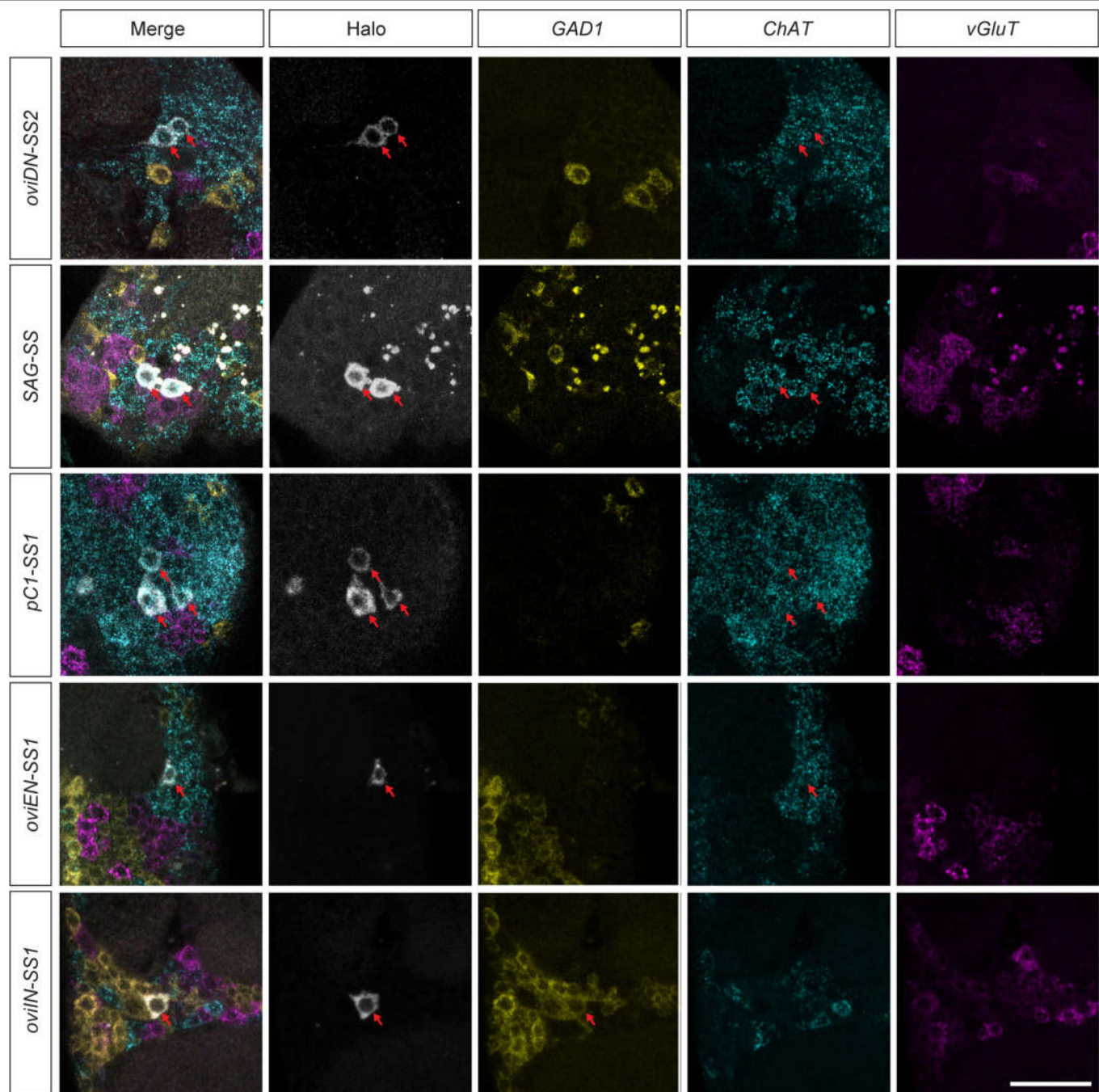


Extended Data Fig. 1 | Split-GAL4 driver lines targeting *oviDNs*, *SPSNs*, *SAG* neurons, *pC1* neurons, *oviENs* and *oviINs*. Confocal images of the central nervous system from female and male flies carrying the indicated split-GAL4 driver lines as well as *UAS-myrFLAG* or *UAS-CsChrimson-mVenus*. Samples were stained with the monoclonal antibody nc82 to reveal all synapses (magenta),

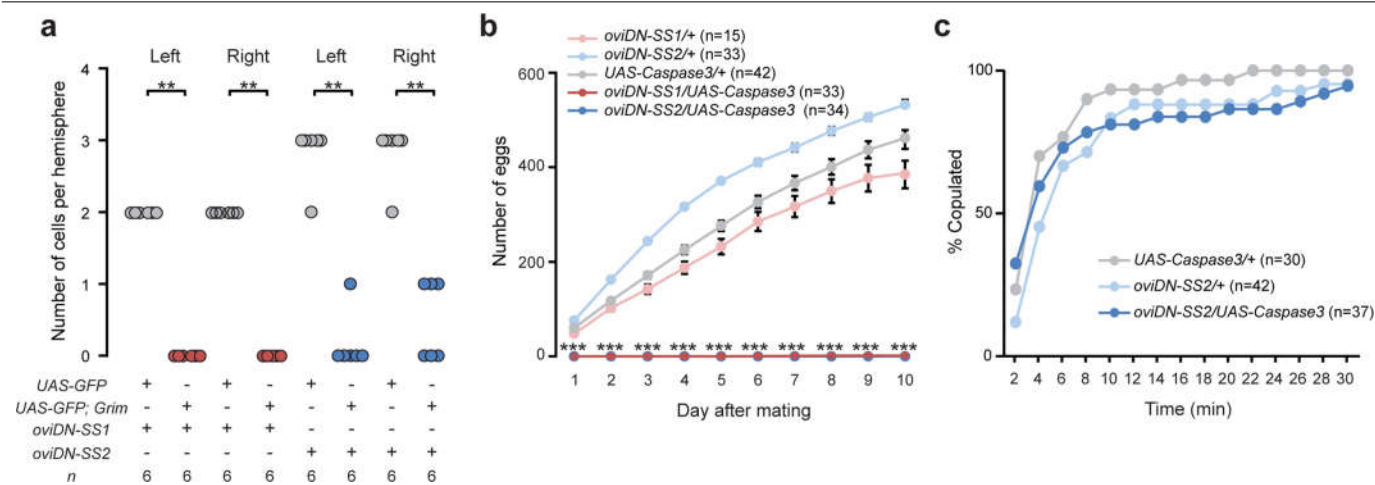
and with anti-FLAG or anti-GFP to reveal the membranes of targeted neurons (green). Scale bar, 100 μ m. Both *oviDN-SS1* and *oviDN-SS2* label a single *oviDNa* and a single *oviDNb* cell in each hemisphere; *oviDN-SS2* also weakly labels an unrelated cell (pMP1) that is present in both sexes.



Extended Data Fig. 2 | Expression of *fru* and *dsx* in *oviDN*s and *pC1* neurons. Confocal images of female brains showing the co-labelling of *oviDN*-SS lines with *fru*-LexA but not *dsx*-LexA, and of the *pC1*-SS1 line with *dsx*-LexA but not *fru*-LexA. Scale bars, 20 μ m.

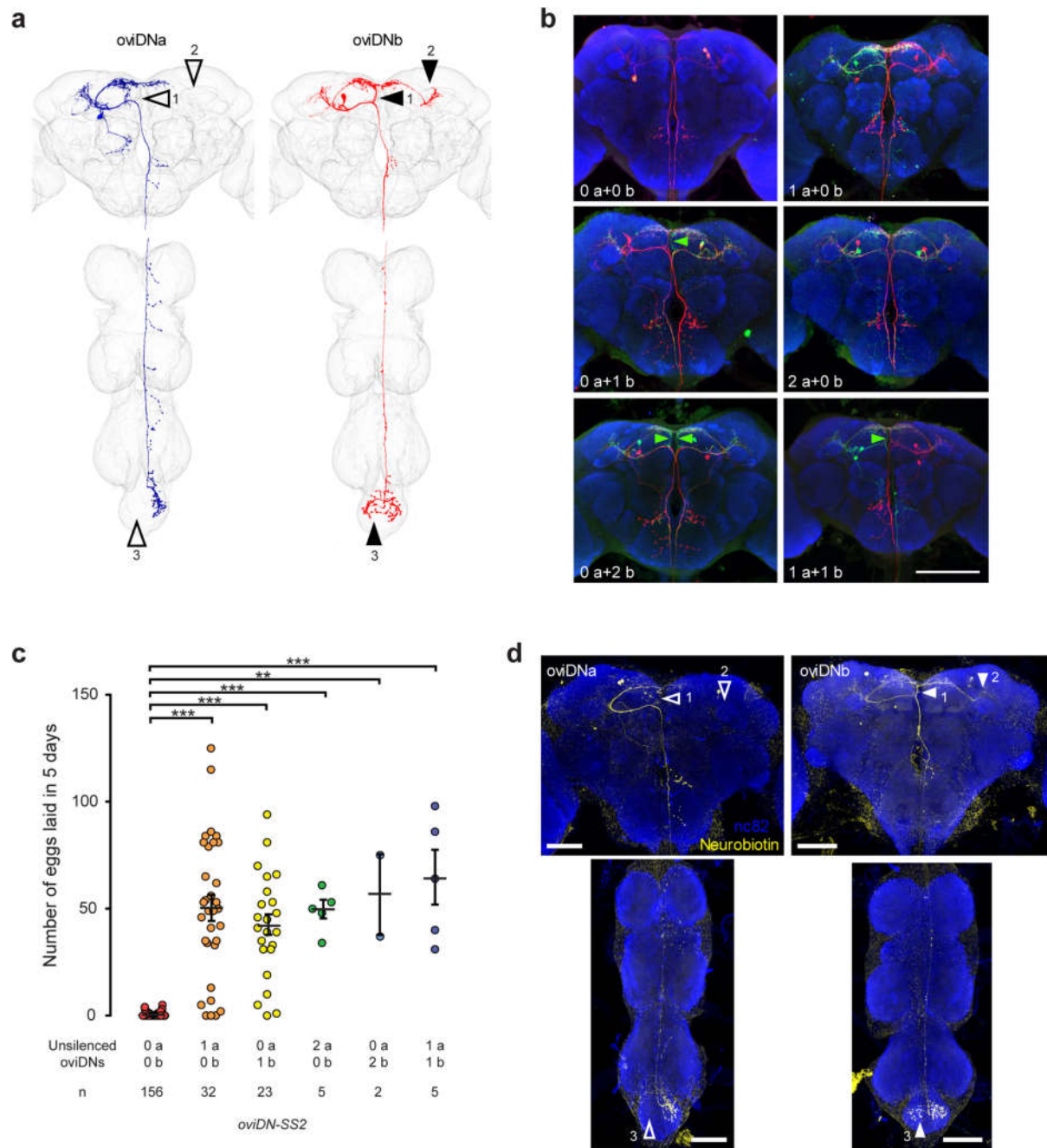


Extended Data Fig. 3 | Neurotransmitter types revealed by fluorescence in situ hybridization. Confocal images showing the expression of *GAD1*, *ChAT* and *vGluT* in *oviDN*s, *SAG* neurons, *pC1* neurons, *oviEN*s and *oviIN*s in female brains. Red arrows indicate cell bodies of interest. Scale bars, 20 μ m.



Extended Data Fig. 4 | oviDNs are required for oviposition but not copulation. **a**, Number of GFP-expressing neurons in female brains of the indicated genotypes. ****** $P < 0.01$ by Wilcoxon test. **b**, Total number of eggs laid by individual mated females over 10 consecutive days after mating, shown as

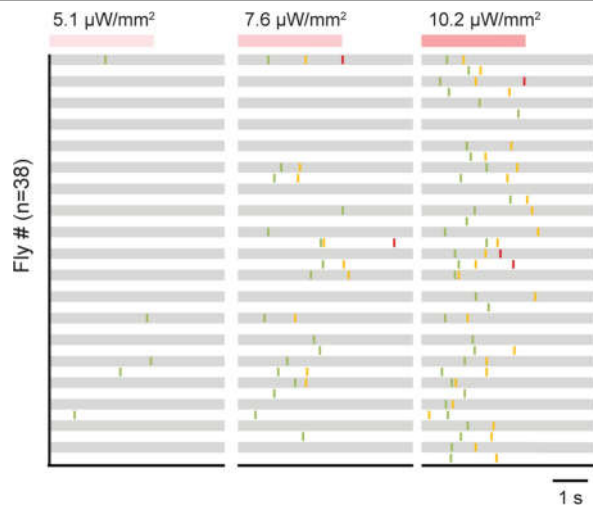
mean \pm s.e.m. Note the values of zero for both oviDN-ablated genotypes at all time points. ******* $P < 0.001$ by Wilcoxon test. **c**, Cumulative traces showing the percentage of females copulating over a 30-min observation period.



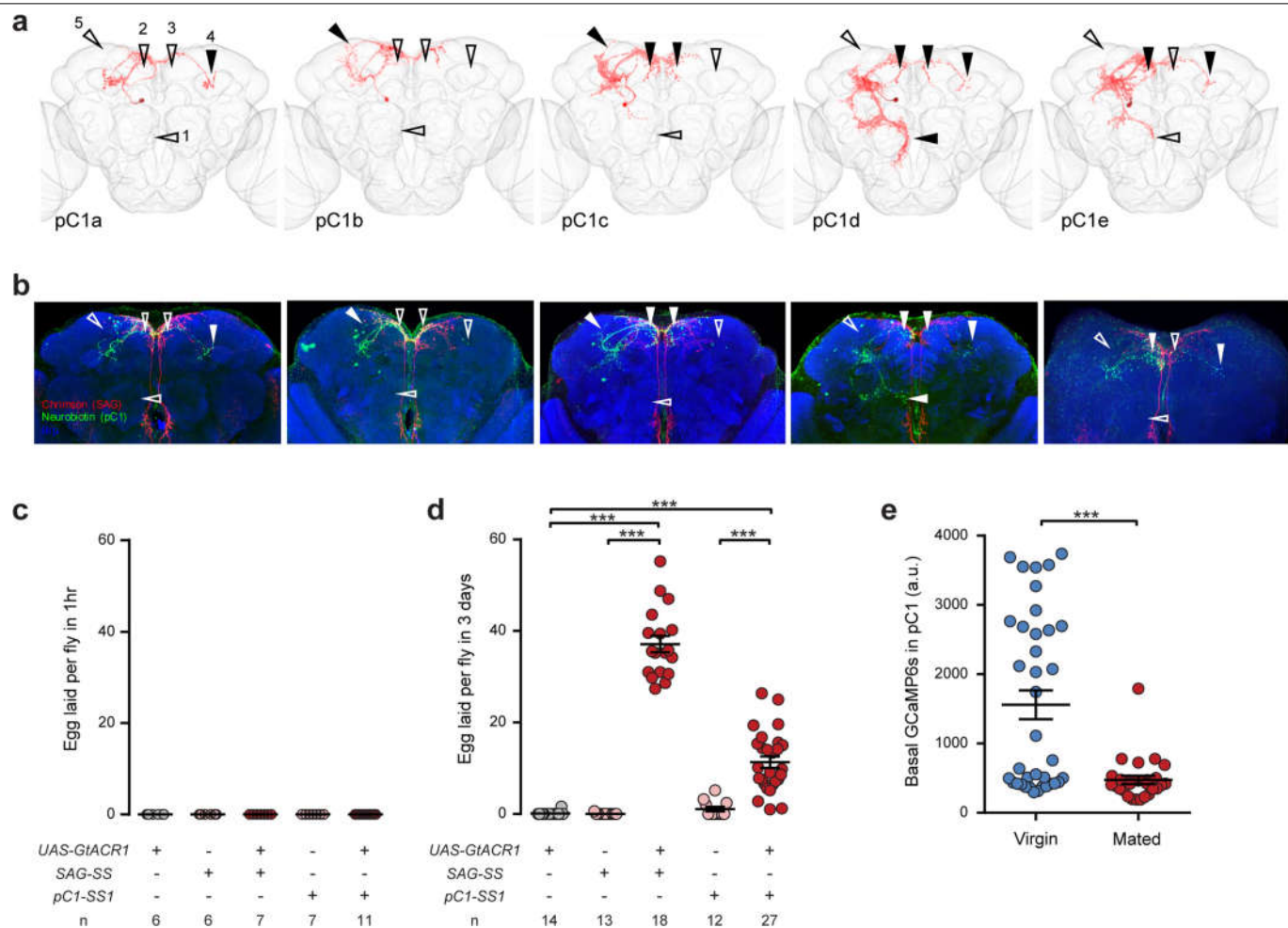
Extended Data Fig. 5 | Stochastic labelling and unsilencing of oviDNs.

a, Images of two female samples in which a single oviDNA or oviDNb cell is labelled, as shown in Fig. 1b. Arrowheads indicate branches that are present in oviDNb (solid) but absent in oviDNA (open). The branch that is labelled by arrowhead 1 was primarily used to distinguish oviDNA from oviDNb. **b**, Example images of brains in which oviDNs were either silenced (red; *Kir2.1::tdTomato*) or unsilenced (green; *mCD8::GFP*). The number of unsilenced oviDNs in each sample is shown. Green arrowheads indicate distinctive branches of oviDNb.

Brains were counterstained with nc82 (blue). Scale bar, 100 μ m. **c**, Number of eggs laid in the five days after mating by mated females with different oviDNs unsilenced. *** P < 0.001, ** P < 0.01 by Wilcoxon test. Scatter plots show mean \pm s.e.m. **d**, Confocal images of two samples in which a single oviDN was loaded with neurobiotin during whole-cell recording. The samples were stained with streptavidin (to reveal the recorded cell, yellow) and nc82 (blue). Arrowheads indicate oviDNb-specific branches. Scale bars, 100 μ m.

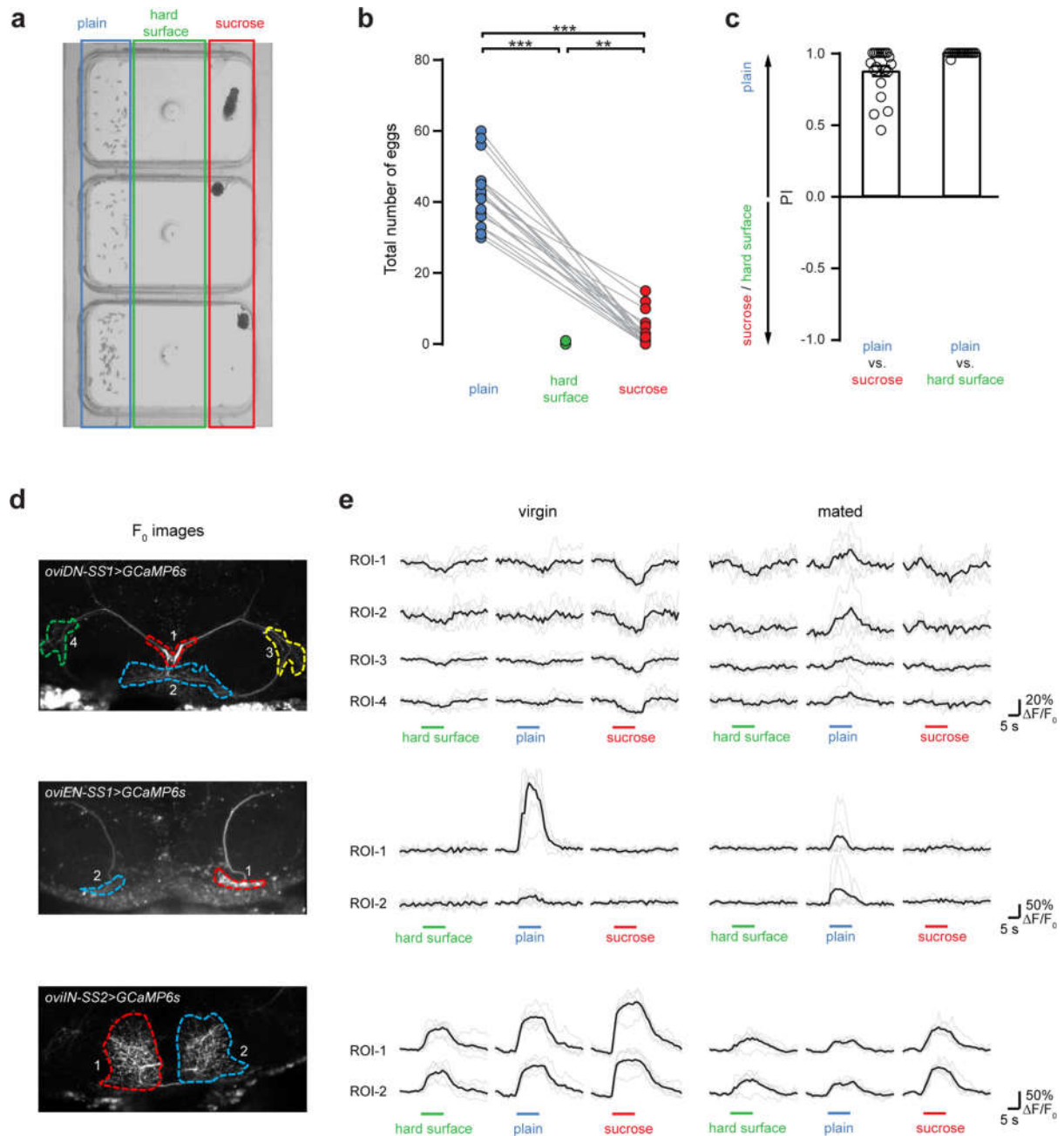


Extended Data Fig. 6 | Sequence of oviposition actions after oviDN stimulation. Example ethograms showing the onsets of oviposition actions in mated females after photoactivation (3 s) of oviDNs at varying light intensities. Each row represents a single female.



Extended Data Fig. 7 | Anatomical and functional characterization of pC1 neurons. **a.** Confocal images of single pC1 neurons in the female brain, as shown in Fig. 3d. Arrowheads indicate the presence (solid) or absence (open) of subtype-specific branches. **b.** Confocal images of neurobiotin-filled pC1 neurons from which whole-cell patch recordings were obtained, indicating the

branches that were used for subtype identification as in **a. c, d**, Number of eggs laid by virgin females during a one-hour (**c**) or three-day (**d**) period in which either SAG or pC1 neurons were optogenetically silenced. **e.** Basal GCaMP6s signals in pC1 cell bodies in virgin and mated females. *** $P < 0.001$ by Wilcoxon test; scatter plots show mean \pm s.e.m. (**d, e**).



Extended Data Fig. 8 | Egg-laying substrate preferences and substrate-evoked calcium responses in *oviDNs*, *oviENs* and *oviINs*. **a**, Image of the egg-laying chambers in each of which an individual mated female had laid numerous eggs. Chambers with plain agarose (blue box), agarose containing 150 mM sucrose (red box) and plastic surface (green box) are indicated. **b**, Total number of eggs laid by individual mated females in a 12-h observation period. $**P < 0.01$, $***P < 0.001$ by Wilcoxon test. **c**, Preference indices showing the preference of female flies for laying eggs on different substrates. Preference index (PI) is

calculated as (number of eggs on plain agarose – number of eggs on other substrate)/total number of eggs. Data are mean \pm s.e.m. **d**, Projected images of *oviDNs* (top), *oviENs* (middle) and *oviINs* (bottom) expressing GCaMP6s, showing ROIs for quantification. **e**, Example $\Delta F/F_0$ traces for each ROI upon presentation of the indicated substrates, in virgin (left) and mated (right) females. Horizontal bars indicate presentation of the substrate. Darker traces are averaged from six trials (lighter traces).

Extended Data Table 1 | Synaptic connections identified by electron microscopy reconstruction

Pre	Cell ID	Post											
		SAG_R	SAG_L	pC1a	pC1b	pC1c	pC1d	pC1e	oviIN	oviEN	oviDNa	oviDNa	oviDNb
SAG_R [#]	5353954	0	0	173	17	78	0	2	0	0	0	0	1
SAG_L [#]	4358525	0	0	79	2	19	0	0	0	0	0	0	0
pC1a [*]	3807213	4	1	3	40	85	20	7	0	0	0	0	1
pC1b [#]	3781622	0	0	0	0	6	0	0	0	0	0	0	0
pC1c [*]	3794184	0	0	5	6	0	10	23	2	0	1	0	0
pC1d [#]	3778246	0	0	2	0	5	0	10	20	0	1	0	0
pC1e [*]	1269969	0	0	1	3	4	5	4	31	0	0	0	0
oviIN [#]	6244095	0	0	0	0	75	0	78	0	176	20	35	29
oviEN [#]	1259227	0	0	2	0	0	0	6	136	0	118	42	142
oviDNa [#]	5143347	0	0	0	0	1	0	1	0	0	0	0	0
oviDNa [#]	1875105	0	0	0	0	0	0	0	0	0	0	1	0
oviDNb [#]	1862763	0	0	0	0	0	0	0	0	0	0	0	0

*Fully traced cells; #partially traced cells. SAG_R and SAG_L indicate right- and left-hemisphere SAG cells, respectively; all other neurons are right-hemisphere cells.

Extended Data Table 2 | oviDN inputs identified by electron microscopy reconstruction

Cell ID	Hemisphere	Cell type	oviDNa (5143347)	oviDNa (1875105)	oviDNb (1862763)	Total
1259227	R	oviEN	118	42	142	302
6244095	R	oviIN	20	35	29	84
3353966	R		31	1	27	59
4634382	R		3	38	16	57
2361058	R		19	10	15	44
1334539	L		21	2	13	36
5870279	L	oviIN	6	10	11	27
5390561	R		8	0	16	24
4590002	N.D.		16	4	3	23
1879478	R		0	21	1	22
2141316	R		4	11	7	22
11122221	R		13	1	8	22
2360875	R		1	17	0	18
8460445	L		13	0	4	17
2712415	R		2	13	1	16
3243035	R		7	0	9	16
5330678	L		13	0	3	16
4295394	L		9	0	6	15
7054780	N.D.		9	0	6	15
2232454	R		0	14	0	14
2613258	R		0	13	1	14
3188249	R		8	2	4	14
1576051	R		0	0	13	13
5470288	N.D.		9	0	4	13
3588146	R		5	2	5	12
4344860	R		5	1	5	11
5316770	R		10	1	0	11
5325544	R		10	0	1	11
5431073	L		5	0	6	11
6759088	L		8	0	3	11
7021239	L		0	11	0	11
7532739	N.D.		4	0	7	11
2255653	R		6	0	4	10
3709065	N.D.		0	8	2	10
9040679	L	oviEN	0	8	2	10

Number of synaptic connections identified between various input neurons and the right-hemisphere oviDNa and oviDNb cells. R and L indicate the soma location in the right (ipsilateral) or left (contralateral) hemisphere; N.D., soma not identified.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection pCLAMP 10 from Molecular Devices was used to collect electrophysiology data.

Data analysis pCLAMP 10 from Molecular Devices and MATLAB from Mathworks were used to analyze electrophysiology data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples sizes were chosen a priori according to typical sample sizes in the field
Data exclusions	No data were excluded.
Replication	All attempts at replication were successful.
Randomization	Samples were allocated based on the corresponding genotypes.
Blinding	Investigators were not blinded to group allocation during data collection and analysis, as these were nonsubjective.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Antibodies used in this study are described in Methods section.
Validation	Validation statement of antibodies used in this study are available on the manufacturers' websites.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Drosophila melanogaster were used in this study. Detailed information are provided in Extended Data.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	No ethical approval was required.

Note that full information on the approval of the study protocol must also be provided in the manuscript.


Caveolae in CNS arterioles mediate neurovascular coupling

<https://doi.org/10.1038/s41586-020-2026-1>

Received: 19 February 2019

Accepted: 3 January 2020

Published online: 19 February 2020

 Check for updates

Brian W. Chow^{1,3}, Vicente Nuñez^{1,3}, Luke Kaplan¹, Adam J. Granger^{1,2}, Karina Bistrong¹, Hannah L. Zucker¹, Payal Kumar¹, Bernardo L. Sabatini^{1,2} & Chenghua Gu^{1✉}

Proper brain function depends on neurovascular coupling: neural activity rapidly increases local blood flow to meet moment-to-moment changes in regional brain energy demand¹. Neurovascular coupling is the basis for functional brain imaging², and impaired neurovascular coupling is implicated in neurodegeneration¹. The underlying molecular and cellular mechanisms of neurovascular coupling remain poorly understood. The conventional view is that neurons or astrocytes release vasodilatory factors that act directly on smooth muscle cells (SMCs) to induce arterial dilation and increase local blood flow¹. Here, using two-photon microscopy to image neural activity and vascular dynamics simultaneously in the barrel cortex of awake mice under whisker stimulation, we found that arteriolar endothelial cells (aECs) have an active role in mediating neurovascular coupling. We found that aECs, unlike other vascular segments of endothelial cells in the central nervous system, have abundant caveolae. Acute genetic perturbations that eliminated caveolae in aECs, but not in neighbouring SMCs, impaired neurovascular coupling. Notably, caveolae function in aECs is independent of the endothelial NO synthase (eNOS)-mediated NO pathway. Ablation of both caveolae and eNOS completely abolished neurovascular coupling, whereas the single mutants exhibited partial impairment, revealing that the caveolae-mediated pathway in aECs is a major contributor to neurovascular coupling. Our findings indicate that vasodilation is largely mediated by endothelial cells that actively relay signals from the central nervous system to SMCs via a caveolae-dependent pathway.

Despite representing only 2% of body mass, the brain uses 20% of the body's energy at rest and has very limited ability to store energy³. To meet acute changes in regional brain energy demand, a process called neurovascular coupling rapidly increases local blood flow following neural activation¹; this is also the basis for functional brain imaging, one of the few techniques currently available to image and measure activity in the human brain in both health and disease².

Neurovascular coupling begins with increased neural activity and ends with SMC relaxation leading to arteriolar vasodilation and increased capillary blood flow^{2,4}. This process occurs rapidly, on the order of hundreds of milliseconds *in vivo* under physiological conditions^{5–7}. How signals are transmitted from neurons to SMCs is not completely understood. The conventional view has been that following neural activity, neurons and astrocytes release vasodilatory signals that act directly on SMCs to relax and expand arteriolar diameter to increase blood flow¹. However, recent studies have indicated that blood vessels can also sense changes in neural activity^{6,8,9}, but the mechanisms underlying how these endothelial cells (ECs) in the central nervous system (CNS) mediate neurovascular coupling remains largely unknown.

Here, we demonstrate that CNS aECs actively mediate signals from neurons to facilitate the relaxation of SMCs during neurovascular coupling. We found that unlike other segments of ECs in the CNS vasculature, aECs contain abundant caveolae. We used *in vivo* two-photon

microscopy for simultaneous measurement of neural activity and vascular dynamics (arteriolar vessel diameter and capillary blood flow) in the barrel cortex of awake mice following whisker stimulation. Acute genetic perturbations that eliminated caveolae in aECs, but not in neighbouring SMCs, impaired neurovascular coupling. Moreover, caveolae function in aECs is independent of the eNOS-mediated NO pathway, and ablation of both caveolae and eNOS completely abolished neurovascular coupling, revealing that the caveolae-mediated pathway in aECs is a major contributor to neurovascular coupling. Finally, we demonstrate that MFSD2A, a molecular suppressor of caveolae formation, is absent in aECs and that ectopic expression of MFSD2A specifically in aECs is sufficient to impair neurovascular coupling. Our findings indicate that ECs actively mediate signals from neurons and astrocytes to SMCs via a caveolae-dependent pathway, which is a major mechanism underlying neurovascular coupling.

CNS arteriolar ECs have abundant caveolae

It has previously been shown that caveolae are actively suppressed in most CNS ECs to ensure blood–brain barrier integrity^{10–12}. However, this suppression of caveolae is not uniform in all segments of the CNS endothelium as aECs have abundant caveolae (Fig. 1b, d), in contrast to negligible numbers of caveolae in capillary ECs (cECs) (Fig. 1a, d).

¹Department of Neurobiology, Harvard Medical School, Boston, MA, USA. ²Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, MA, USA. ³These authors contributed equally: Brian W. Chow, Vicente Nuñez. ✉e-mail: chenghua_gu@hms.harvard.edu

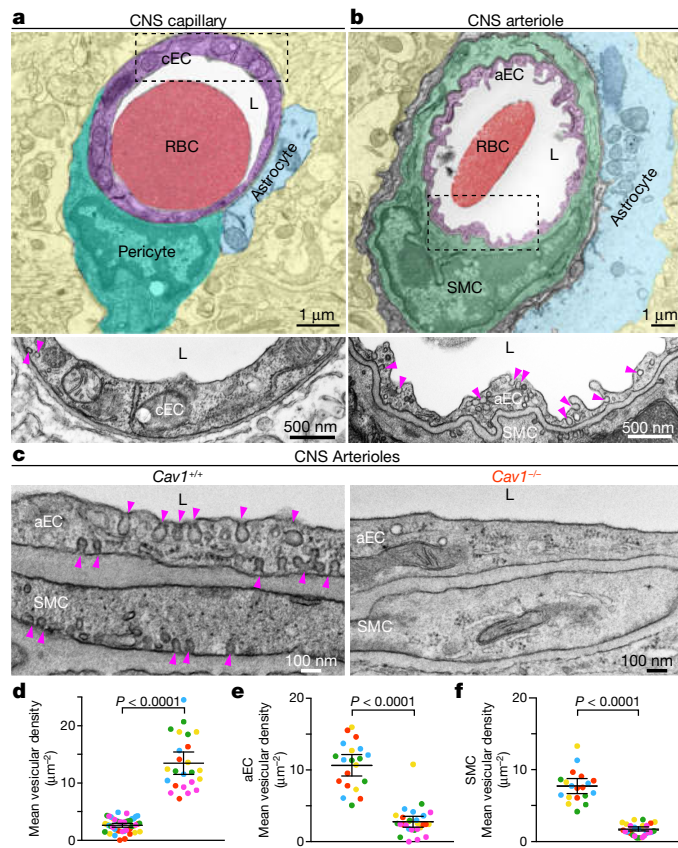


Fig. 1 | CNS arterioles have abundant caveolae. **a**, Transmission electron microscopy image of a CNS capillary. Pseudocolours highlight different cells: cEC (purple), pericyte (teal), astrocyte end-foot (blue), red blood cell (RBC, red), lumen (L) (white) and neuropil (yellow). Bottom shows an inverted, magnified image of the boxed area in the top panel. **b**, Transmission electron microscopy image of a CNS arteriole. Pseudocolours highlight different cells: aEC (purple), SMC (green), astrocyte end-foot (blue) and neuropil (yellow). Bottom shows a magnified image of the boxed area in the top panel. Arrowheads point to vesicles (**a**, **b**). **c**, Transmission electron microscopy images of aECs and SMCs from *Cav1*^{+/+} and *Cav1*^{-/-} mice. Arrowheads point to caveolae. **d**, Mean vesicular density in cECs and aECs from wild-type mice (*n* = 5 mice, 46 capillaries and 24 arterioles). **e**, **f**, Mean vesicular density in aECs (**e**) and SMCs (**f**) in *Cav1*^{+/+} (*n* = 5 mice, 20 arterioles) and *Cav1*^{-/-} mice (*n* = 5 mice, 28 arterioles). Data are mean ± s.e.m.; nested, unpaired, two-tailed *t*-test (**d–f**).

cECs and aECs can be distinguished under transmission electron microscopy: capillaries have a smooth lumen whereas arterioles have a ruffled lumen; cECs are also surrounded by pericytes, whereas aECs are sheathed by SMCs (Fig. 1a, b). The abundant vesicles in aECs are abolished in *caveolin-1*-mutant (*Cav1*^{-/-}) mice, suggesting that they are composed of caveolae (Fig. 1c, e). Caveolin 1 is an essential component of caveolae and the endothelium of *Cav1*^{-/-} mice lacks caveolae^{11–14}. Thus, CNS aECs have abundant caveolae, consistent with a previous study¹⁵. Notably, many caveolae are also present in the SMCs that wrap around the aECs (Fig. 1c, f).

Neurovascular coupling requires caveolae

Because caveolae are specifically abundant in aECs, and CNS arterioles are the site of vasodilation^{2,4}, we examined *Cav1*^{-/-} mice to determine whether caveolae are important for neurovascular coupling. To study neurovascular coupling in vivo, we optimized a two-photon microscope for simultaneous measurement of neural activity and vascular dynamics, including arteriolar vessel diameter and capillary blood flow at single-vessel resolution in awake mice (Extended Data Fig. 1).

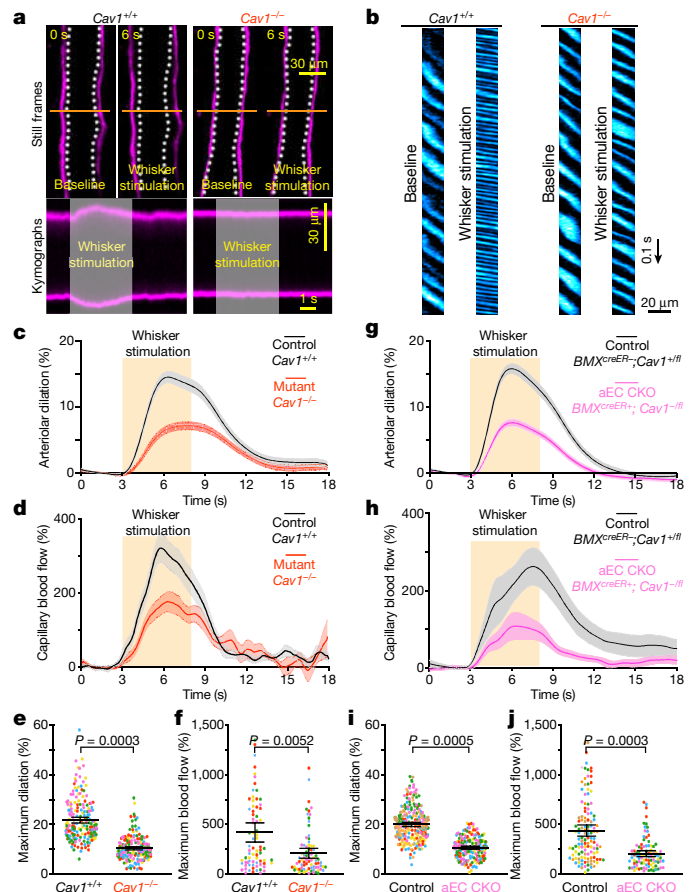


Fig. 2 | Caveolae in CNS aECs specifically are required for neurovascular coupling. **a**, Still frame images of pial arteries during neurovascular coupling in *Cav1*^{+/+} and *Cav1*^{-/-} mice using in vivo two-photon microscopy. Top, Hydrizide-stained arterioles during baseline and whisker stimulation. White hashes outline the arterioles during baseline period. Bottom, kymographs of the arteriolar dilation, generated by transverse line scans (orange lines in top images). The grey rectangle in the kymograph represents the whisker stimulation period. **b**, Kymographs of red blood cell flow in capillaries for *Cav1*^{+/+} and *Cav1*^{-/-} mice. Dark streaks represent red blood cells, blue streaks represent the fluorescent tracer-filled capillary lumen. Left and right kymographs show red blood cell flow during baseline and whisker stimulation, respectively. **c–f**, Time course of change in arteriolar dilation (**c**), change in red blood cell velocity (**d**), maximum change in arteriolar dilation (**e**) and maximum red blood cell velocity (**f**) in *Cav1*^{+/+} (*n* = 5 mice, 196 arterioles, 77 capillaries) and *Cav1*^{-/-} mice (*n* = 5 mice, 194 arterioles, 79 capillaries). **g–j**, Time course of change in arteriolar dilation (**g**), change in red blood cell velocity (**h**), maximum change in arteriolar dilation (**i**) and maximum red blood cell velocity (**j**) in control (*BMXcreER-/-; Cav1*^{+/+}; *n* = 7 mice, 260 arterioles, 122 capillaries) and aEC conditional *Cav1*-knockout mice (*BMXcreER-/-; Cav1*^{-/-}; *n* = 5 mice, 193 arterioles, 94 capillaries). Data are mean ± s.e.m.; nested, unpaired, two-tailed *t*-test (**e**, **f**, **i**, **j**).

We focused on the barrel cortex, a well-characterized region of mouse somatosensory cortex that processes sensory input from the vibrissae¹⁶. Sensory stimulation by whisker brushing in awake mice evoked spatially and temporally patterned neural activity that can be imaged in the barrel cortex by intracellular calcium levels in mice expressing the calcium sensor GCaMP6s in neurons (*Thy1*-GCaMP6s) (Extended Data Fig. 1b, c). Hydrizide⁵ and quantum dots were injected intravenously into *Thy1*-GCaMP6s mice to visualize arterioles and image capillary blood flow, respectively. Upon whisker brushing, we observed a robust increase in the GCaMP signal in neurons, followed by arteriolar dilation and increased red blood cell velocity (measured by tracking the movement of red blood cells, which are devoid of quantum dots and thus appear dark) (Extended Data Fig. 1b–g, Supplementary Videos 1–3).

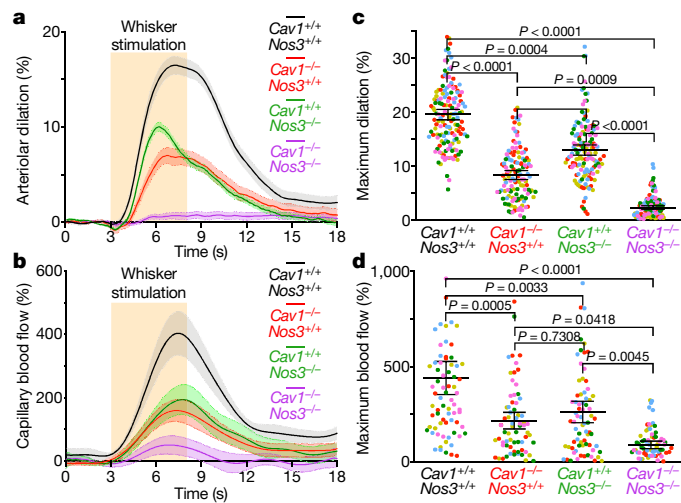


Fig. 3 | Caveolae in aECs mediate neurovascular coupling independently of eNOS. Time course of change in arteriolar dilation (a), change in red blood cell velocity (b), maximum percentage change in arteriolar dilation (c) and maximum percentage change in red blood cell velocity (d) in *Cav1*^{+/+}*Nos3*^{+/+} (*n* = 5 mice, 148 arterioles, 76 capillaries), *Cav1*^{-/-}*Nos3*^{+/+} (*n* = 5 mice, 128 arterioles, 68 capillaries), *Cav1*^{+/+}*Nos3*^{-/-} (*n* = 5 mice, 137 arterioles, 73 capillaries) and *Cav1*^{-/-}*Nos3*^{-/-} mice (*n* = 5 mice, 139 arterioles, 74 capillaries). Data are mean \pm s.e.m.; nested, one-way analysis of variance (ANOVA) with a post hoc Bonferroni multiple comparison adjustment (c, d).

Finally, in contrast to the robust vasodilation observed in the barrel cortex (using our *in vivo* whisker stimulation paradigm), the retrosplenial cortex—a brain region not associated with processing whisking¹⁷—exhibits very low levels of vasodilation (Extended Data Fig. 1h, i). This result indicates that the changes in arterial vessel diameter are a result of the whisker-stimulus-dependent neural activity and not systemic variables.

Cav1^{-/-} mice exhibited attenuated arteriolar dilation upon whisker stimulation, whereas arterioles from wild-type *Cav1*^{+/+} and heterozygous *Cav1*^{+/-} mice dilated robustly (Fig. 2a, c, e, Extended Data Fig. 3a, i, Supplementary Video 4). Moreover, this vasodilation defect was observed in both pial arteries and penetrating arterioles diving deep into the parenchyma in *Cav1*^{-/-} mice compared with their wild-type littermates (Extended Data Fig. 2). Notably, the baseline diameter and latency to dilate were similar across the three genotypes (Extended Data Fig. 3b–d). These results suggest that the absence of caveolae does not impair basal vessel tone and kinetics but specifically impairs the amplitude of sensory-evoked arteriolar dilation. Consistent with the attenuation of arteriolar vasodilation, capillary blood flow was also impaired in mutant *Cav1*^{-/-} mice upon whisker stimulation compared to control mice (Fig. 2b, d, f, Supplementary Video 5), whereas baseline capillary velocity and kinetics were similar across genotypes (Extended Data Fig. 3e, f). Moreover, the attenuated arteriolar dilation and capillary blood flow in mutant *Cav1*^{-/-} mice were not due to either the impairment in sensory-evoked neural activity or alteration in blood pressure, because control and mutant mice display similar GCaMP6s dynamics in neurons (Extended Data Fig. 3g, h) and similar systolic, diastolic and mean blood pressure (Extended Data Fig. 3j). The normal blood pressure observed in *Cav1*^{-/-} mutant mice is consistent with previous studies¹⁸. Thus, these results demonstrate that caveolae are essential for optimal neurovascular coupling.

Because SMCs control arteriolar dilation during neurovascular coupling²⁴, we next examined whether the attenuated neurovascular coupling in *Cav1*^{-/-} mice is due to impaired integrity and function of SMCs. To visualize SMC morphology and vessel coverage, we intravenously injected hyaluronidase into control *Cav1*^{+/+};NG2-DsRed⁺ and mutant *Cav1*^{-/-};NG2-DsRed⁺ mice. NG2-DsRed is a reporter for SMCs,

oligodendrocytes and pericytes^{16,19}. By quantifying of the number of DsRed⁺ cells on hyaluronidase⁺ arterioles, we established that there is normal coverage and morphology of SMCs in *Cav1*^{-/-} mice compared with wild type (Extended Data Fig. 4a, b). Moreover, we found similar expression of various contractile proteins, including α -smooth muscle actin (SMA), MYH11, transgelin and desmin in SMCs in *Cav1*^{-/-} mice and wild-type littermates (Extended Data Fig. 4c–g).

To examine whether ablation of caveolae affects the ability of SMCs to respond to contractile and vasodilatory signals, we imaged arteriolar diameter changes in acute brain slices under two-photon microscopy after delivery of contractile and vasodilatory pharmacological compounds. We found that SMCs in *Cav1*^{-/-} mice displayed normal contraction compared with wild-type controls following administration of U46619, a thromboxane A2 receptor agonist (Extended Data Fig. 4h, i, Supplementary Videos 6, 7). When diethylamine (DEA)-NONOate, a NO donor, was subsequently applied to the same vessel, we observed a similar level of dilation as in the wild-type controls (Extended Data Fig. 4h, j, Supplementary Videos 6, 7). Finally, to examine whether the impaired neurovascular coupling in mutant *Cav1*^{-/-} mice is due to the inability of SMCs to relax following release of vasodilatory signals *in vivo*, we used two-photon microscopy in anesthetized mice and imaged changes in vasodilation upon superfusing DEA NONOate onto the pia of control *Cav1*^{+/+} and mutant *Cav1*^{-/-} mice. Upon acute administration of DEA NONOate, the SMCs from both control *Cav1*^{+/+} and mutant *Cav1*^{-/-} mice relax and dilate arterioles at similar levels *in vivo* (Extended Data Fig. 4k, l, Supplementary Videos 8, 9). These experiments demonstrate that the absence of caveolae impairs neurovascular coupling despite the presence of functionally normal SMCs.

Neurovascular coupling requires aEC caveolae

Caveolae are present in both aECs and SMCs in the CNS (Fig. 1c, f). We next tested whether caveolae function in a cell-autonomous manner using acute, cell-type-specific deletion of *Cav1* in adult mice. First, we crossed *BMX*^{creER} mice, a tamoxifen-inducible aEC-specific driver line²⁰, with *Cav1*-floxed mice (*Cav1*^{fl/fl})²¹ to acutely ablate caveolae only in aECs. After tamoxifen treatment, CAV1 was specifically lost in aECs but not in SMCs of *BMX*^{creER}; *Cav1*^{fl/fl} mutant mice; CAV1 protein was present in both aECs and SMCs in *BMX*^{creER}; *Cav1*^{+/+} control mice. (Extended Data Fig. 5a, c). Transmission electron microscopy analysis showed that in the barrel cortex of mutant *BMX*^{creER}; *Cav1*^{fl/fl} mice, caveolae were ablated acutely in aECs but were still present in SMCs, whereas abundant caveolae were present in both aECs and SMCs in the CNS of control *BMX*^{creER}; *Cav1*^{+/+} mice (Extended Data Fig. 5b, d). Using our *in vivo* imaging paradigm, we found attenuation in arteriolar dilation (Fig. 2g, i, Extended Data Fig. 6a) and capillary blood flow (Fig. 2h, j) in *BMX*^{creER}; *Cav1*^{fl/fl} mice on whisker stimulation, similar to observations in the *Cav1*^{-/-} mutant mice (Fig. 2a–f). Baseline and kinetics of arteriolar diameter and capillary blood flow were unaffected in tamoxifen-treated *BMX*^{creER}; *Cav1*^{fl/fl} mice (Extended Data Figs. 5e, f, 6b–d). Together, these experiments demonstrate that caveolae in aECs are important for neurovascular coupling.

SMC caveolae are dispensable

To test the role of SMC caveolae in neurovascular coupling, we crossed *Myh11*^{creER} mice, a tamoxifen-inducible SMC driver line²², to *Cav1*-floxed mice to acutely ablate caveolae in SMCs. After tamoxifen treatment, we found that both CAV1 protein and caveolae were ablated successfully in SMCs but preserved in aECs in *Myh11*^{creER}; *Cav1*^{fl/fl} mutant mice (Extended Data Fig. 7a–d). However, in contrast to the attenuated neurovascular coupling observed in mice lacking CAV1 in aECs, no impairment in arteriolar dilation and capillary blood flow after whisker stimulation were observed in mutants lacking CAV1 and caveolae in SMCs (Extended Data Figs. 6e–h, 7e–j), indicating that caveolae in SMCs have a negligible role—if any—in neurovascular coupling.

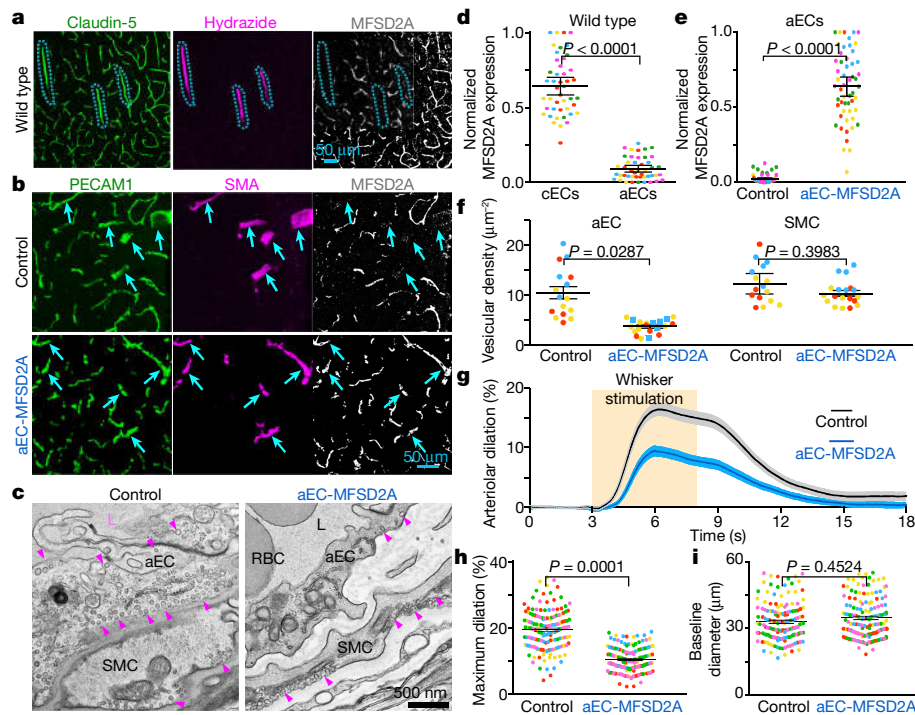


Fig. 4 | CNS arterioles do not express MFSD2A and ectopic expression of MFSD2A in aECs downregulates caveolae and attenuates neurovascular coupling. **a**, Immunostaining for claudin 5, hydrizide and MFSD2A on wild-type adult brain sections demonstrates that MFSD2A is not detectable in CNS arterioles. Blue hashes outline the hydrizide⁺ arterioles. **b**, Immunostaining on adult brain sections for PECAM1, SMA and MFSD2A from control (*BMX^{creER};R26^{LSL-Mfsd2a}/+*) mice and mice overexpressing MFSD2A specifically in the aEC (*BMX^{creER};R26^{LSL-Mfsd2a}/+*). Arrowheads point to SMA⁺ arteries. **d**, Quantification of the normalized immunofluorescence of MFSD2A in cECs (hydrizide⁺ claudin 5⁺) and aECs (hydrizide⁺ claudin 5⁺) ($n = 5$ mice, 45 images) as shown in **a**, **e**, Normalized MFSD2A immunofluorescence in aECs

Caveolae in aECs function independently of eNOS

Next, we examined how aECs utilize caveolae to mediate neurovascular coupling. Caveolae have been implicated in many cellular processes²³, including transcytosis^{11,12}, serving as a membrane reservoir during mechanical stretch²⁴, clustering receptors and ion channels²⁵, and mediating intracellular signalling²⁶. We focused on NO signalling because NO is a major vasodilatory factor in neurovascular coupling²⁷ and previous studies have reported that CAV1 interacts physically with eNOS (encoded by *Nos3*)²⁸. We first examined whether eNOS and NO levels are altered in the absence of caveolae. Unexpectedly, we found similar levels of eNOS protein and NO in aECs in wild-type control and *Cav1*^{-/-} mice, whereas both eNOS protein and NO signal were absent in *Nos3*^{-/-} mice (Extended Data Fig. 8a–d). To examine genetic interactions between *Cav1* and *Nos3*, we characterized neurovascular coupling in *Cav1*^{-/-}*Nos3*^{-/-} double-mutant mice. We reasoned that if CAV1 and eNOS are in the same genetic pathway, the double mutants should phenocopy one of the single knockout mice, whereas if they function in separate parallel pathways, the double knockout should have an additive phenotype of the two single-knockout mice. Using our in vivo imaging paradigm, *Nos3*^{-/-} mice displayed attenuated arteriolar dilation and capillary blood flow upon whisker stimulation (Fig. 3a–d), consistent with a previous report²⁷. The *Cav1*^{-/-}*Nos3*^{-/-} double mutant mice completely lost arteriolar dilation and red blood cell velocity enhancement upon whisker stimulation, compared with a partial reduction in *Cav1*^{-/-} and *Nos3*^{-/-} single mutants despite having normal baseline diameter and blood flow (Fig. 3a–d, Extended Data Fig. 8e, f). These results demonstrate that caveolae-mediated neurovascular coupling is independent

from control mice ($n = 4$ mice, 40 images) and mice overexpressing MFSD2A specifically in the aEC ($n = 5$ mice, 51 images) as shown in **b**, **c**, **f**. Transmission electron microscopy images of CNS aECs and SMCs (**c**) and quantification of the mean vesicular density (**f**) in aECs and SMCs from control mice ($n = 3$ mice, 20 arterioles) and mice overexpressing MFSD2A specifically in the aEC ($n = 3$ mice, 22 arterioles). Arrowheads point to caveolae. **g–i**, Time course of change in arteriolar dilation (**g**) and maximum percentage change in arteriolar dilation (**h**) and baseline diameter (**i**) in control mice ($n = 7$ mice, 260 arterioles) and mice overexpressing MFSD2A specifically in the aEC ($n = 5$ mice, 193 arterioles). Data are mean \pm s.e.m.; nested, unpaired, two-tailed *t*-test (**d–f**, **h**, **i**).

of the NO pathway. Moreover, the caveolae-mediated pathway is at least as important as the NO pathway for neurovascular coupling.

MFSD2A downregulates neurovascular coupling

We next investigated why cECs have few caveolae whereas aECs have abundant caveolae. It was previously discovered that MFSD2A expression in CNS cECs actively suppressed caveolae formation and that this was necessary for blood–brain barrier integrity^{10–12}. Using immunohistochemistry, we found that MFSD2A protein was undetectable in aECs in both brain and retina (Fig. 4a, d, Extended Data Fig. 9a–d). Consistent with this result, *Mfsd2a* transcript levels are also low in aECs compared with cECs²⁹. Thus, CNS cECs robustly express *Mfsd2a* to suppress caveolae, whereas aECs lack MFSD2A and are enriched in caveolae. We therefore examined whether ectopic expression of *Mfsd2a* specifically in CNS aECs is sufficient to suppress caveolae in these cells; if so, our results so far predict that this suppression of caveolae in aECs would result in an attenuated neurovascular coupling.

To ectopically express *Mfsd2a* in aECs only, we generated a transgenic mouse in which *Mfsd2a* expression is Cre-dependent (referred to as *R26^{LSL-Mfsd2a}*) (Extended Data Fig. 10a, b) and crossed it to *BMX^{creER}*. After tamoxifen treatment, MFSD2A protein was expressed abundantly in aECs in brains from *BMX^{creER};R26^{LSL-Mfsd2a}/+* mice, whereas control *BMX^{creER};R26^{LSL-Mfsd2a}/+* adult mice lacked MFSD2A expression in brain arterioles (Fig. 4b, e). Moreover, electron microscopy analysis revealed that caveolae density was reduced significantly in *BMX^{creER};R26^{LSL-Mfsd2a}/+* mice relative to control (Fig. 4c, f). Of note, in vivo imaging revealed an attenuation of arteriolar dilation upon whisker stimulation in mice with MFSD2A overexpression in aECs (*BMX^{creER};R26^{LSL-Mfsd2a}/+*) relative

to control mice (*BMX^{creER};R26^{LSL-Mfsd2a}/+*) (Fig. 4g–i, Extended Data Fig. 10e). These experiments demonstrate that ectopic overexpression of MFSD2A in CNS aECs is sufficient to reduce caveolae density and impair neurovascular coupling. Furthermore, inhibition of caveolae vesicles specifically in aECs using two different approaches (overexpression of MFSD2A and genetic deletion of *Cav1*) both resulted in attenuated neurovascular coupling, demonstrating the importance of caveolae in CNS aECs for mediating neurovascular coupling.

Discussion

We used natural stimuli under physiological conditions in awake mice while simultaneously measuring neural activity and vascular dynamics under two-photon microscopy to study mechanisms underlying neurovascular coupling. We discovered that caveolae in CNS aECs have a key role in mediating neurovascular coupling. In addition, we confirmed that the previously reported eNOS pathway also has a role in neurovascular coupling²⁷. However, we found that the caveolae-mediated pathway is independent of eNOS signalling, as the perturbation of both caveolae- and eNOS-mediated pathways together completely abolished neurovascular coupling, whereas ablation of each pathway alone resulted in partial impairment. Thus, these findings indicate that the caveolae-mediated pathway is at least as important as the NO pathway for neurovascular coupling.

Previous studies highlighted the importance of ECs in neurovascular coupling *in vivo*. Locally disrupting ECs using optically induced reactive oxygen species halted propagation of stimulus-evoked vasodilation in pial arteries⁶. Our present findings have extended this work to identify and demonstrate specific molecular and subcellular components in aECs that are essential for neurovascular coupling using cell-type-specific genetic manipulations. Given the recent evidence that cECs are involved in sensing neural activity changes and are important for neurovascular coupling^{2,6,8}, we propose that after sensing nearby increased neural activity, cECs relay this signal electrically to the upstream aECs, which in turn send vasodilatory cues to SMCs via a caveolae-dependent process. We considered how caveolae could carry out this function. Although caveolae have been reported to serve as a membrane reservoir during mechanical stretch²⁴, the attenuated vasodilation observed in *Cav1*^{−/−} mice is unlikely to result from impaired arteriolar elasticity, given that superfusing NO donor *in vivo*—which dilates arterioles by directly relaxing SMC—produced similar dilations in wild-type and *Cav1*^{−/−} mice (Extended Data Fig. 4k, l, Supplementary Videos 8, 9). Similarly, previously described interactions between caveolae and the eNOS-signalling pathway cannot explain the impaired vasodilation observed in *Cav1*^{−/−} mutant mice, because simultaneous ablation of both caveolae- and eNOS-mediated pathways abolishes neurovascular coupling, whereas ablation of either pathway alone results in only partial impairment. In light of these results, the ability of caveolae to cluster ion channels and receptors³⁰—several of which have been implicated in vasodilation⁸—probably explains its role in neurovascular coupling. In aECs, caveolae could cluster these channels to facilitate transmission of vasodilatory signals to SMCs. Identifying the channels that cluster in caveolae in aECs will be an important next step for the field to address.

Our results also demonstrate that ECs from different vascular segments in the CNS exhibit heterogeneity at molecular, cellular and functional levels. Here we show that this endothelial heterogeneity governs the two unique and important functions of the CNS vasculature: the blood–brain barrier and neurovascular coupling. We found that cECs express MFSD2A, which suppresses caveolae to ensure blood–brain-barrier integrity^{10–12}. By contrast, aECs lack MFSD2A and concurrently have abundant caveolae, which are important for neurovascular coupling. We expect this kind of heterogeneity to exist broadly in CNS ECs and that understanding this heterogeneity will advance our understanding of the diverse functions of ECs in health and disease. Given that neurovascular coupling is impaired in various neurological disorders¹, future studies examining whether these different molecular and cellular pathways are altered in disease may provide insight for development of novel therapies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2026-1>.

1. Iadecola, C. The neurovascular unit coming of age: a journey through neurovascular coupling in health and disease. *Neuron* **96**, 17–42 (2017).
2. Hillman, E. M. C. Coupling mechanism and significance of the BOLD signal: a status report. *Annu. Rev. Neurosci.* **37**, 161–181 (2014).
3. Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood–brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150 (2018).
4. Kleinfeld, D. et al. A guide to delineate the logic of neurovascular signaling in the brain. *Front. Neuroenergetics* **3**, 1 (2011).
5. Shen, Z., Lu, Z., Chhatbar, P. Y., O'Herron, P. & Kara, P. An artery-specific fluorescent dye for studying neurovascular coupling. *Nat. Methods* **9**, 273–276 (2012).
6. Chen, B. R., Kozberg, M. G., Bouchard, M. B., Shaik, M. A. & Hillman, E. M. C. A critical role for the vascular endothelium in functional neurovascular coupling in the brain. *J. Am. Heart Assoc.* **3**, e000787 (2014).
7. O'Herron, P. et al. Neural correlates of single-vessel haemodynamic responses *in vivo*. *Nature* **534**, 378–382 (2016).
8. Longden, T. A. et al. Capillary K⁺-sensing initiates retrograde hyperpolarization to increase local cerebral blood flow. *Nat. Neurosci.* **20**, 717–726 (2017).
9. Hogan-Cann, A. D., Lu, P. & Anderson, C. M. Endothelial NMDA receptors mediate activity-dependent brain hemodynamic responses in mice. *Proc. Natl Acad. Sci. USA* **116**, 10229–10231 (2019).
10. Ben-Zvi, A. et al. *Mfsd2a* is critical for the formation and function of the blood–brain barrier. *Nature* **509**, 507–511 (2014).
11. Chow, B. W. & Gu, C. Gradual suppression of transcytosis governs functional blood–retinal barrier formation. *Neuron* **93**, 1325–1333.e3 (2017).
12. Andreone, B. J. et al. Blood–brain barrier permeability is regulated by lipid transport-dependent suppression of caveolae-mediated transcytosis. *Neuron* **94**, 581–594.e5 (2017).
13. Razani, B. et al. Caveolin-1 null mice are viable but show evidence of hyperproliferative and vascular abnormalities. *J. Biol. Chem.* **276**, 38121–38138 (2001).
14. Drab, M. et al. Loss of caveolae, vascular dysfunction, and pulmonary defects in caveolin-1 gene-disrupted mice. *Science* **293**, 2449–2452 (2001).
15. Simionescu, M. et al. The cerebral microvasculature of the rat: structure and luminal surface properties during early development. *J. Submicrosc. Cytol. Pathol.* **20**, 243–261 (1988).
16. Hill, R. A. et al. Regional blood flow in the normal and ischemic brain is controlled by arteriolar smooth muscle cell contractility and not by capillary pericytes. *Neuron* **87**, 95–110 (2015).
17. Lim, D. H. et al. *In vivo* large-scale cortical mapping using channelrhodopsin-2 stimulation in transgenic mice reveals asymmetric and reciprocal relationships between cortical areas. *Front. Neural Circuits* **6**, 11 (2012).
18. Rosengren, B.-I. et al. Transvascular protein transport in mice lacking endothelial caveolae. *Am. J. Physiol. Heart Circ. Physiol.* **291**, H1371–H1377 (2006).
19. Zhu, X., Bergles, D. E. & Nishiyama, A. NG2 cells generate both oligodendrocytes and gray matter astrocytes. *Development* **135**, 145–157 (2008).
20. Ehling, M., Adams, S., Benedito, R. & Adams, R. H. Notch controls retinal blood vessel maturation and quiescence. *Development* **140**, 3051–3061 (2013).
21. Asterholm, I. W., Mundy, D. I., Weng, J., Anderson, R. G. W. & Scherer, P. E. Altered mitochondrial function and metabolic inflexibility associated with loss of caveolin-1. *Cell Metab.* **15**, 171–185 (2012).
22. Wirth, A. et al. G12–G13–LARG-mediated signaling in vascular smooth muscle is required for salt-induced hypertension. *Nat. Med.* **14**, 64–68 (2008).
23. Parton, R. G. Caveolae: structure, function, and relationship to disease. *Annu. Rev. Cell Dev. Biol.* **34**, 111–136 (2018).
24. Sinha, B. et al. Cells respond to mechanical stress by rapid disassembly of caveolae. *Cell* **144**, 402–413 (2011).
25. Balijepalli, R. C. & Kamp, T. J. Caveolae, ion channels and cardiac arrhythmias. *Prog. Biophys. Mol. Biol.* **98**, 149–160 (2008).
26. Sowa, G., Pypaert, M. & Sessa, W. C. Distinction between signaling mechanisms in lipid rafts vs. caveolae. *Proc. Natl Acad. Sci. USA* **98**, 14072–14077 (2001).
27. Toth, P. et al. Purinergic glio-endothelial coupling during neuronal activity: role of P2Y1 receptors and eNOS in functional hyperemia in the mouse somatosensory cortex. *Am. J. Physiol. Heart Circ. Physiol.* **309**, H1837–H1845 (2015).
28. García-Cardeña, G. et al. Dissecting the interaction between nitric oxide synthase (NOS) and caveolin. Functional significance of the nos caveolin binding domain *in vivo*. *J. Biol. Chem.* **272**, 25437–25440 (1997).
29. Vanlandewijck, M. et al. A molecular atlas of cell types and zonation in the brain vasculature. *Nature* **554**, 475–480 (2018).
30. Goedicke-Fritz, S. et al. Evidence for functional and dynamic microcompartmentation of Cav-1/TRPV4/K_{Ca} in caveolae of endothelial cells. *Eur. J. Cell Biol.* **94**, 391–400 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Mice

All mouse experiments were approved by the Harvard University Institutional Animal Care and Use Committee (IACUC). The following mice strains were used: wild type (C57BL/6J, Jackson Laboratory, no. 000664), *Mfsd2a*^{creER} (ref. ³¹), *BMX*^{creER} (ref. ²⁰), *Myh11*^{creER} (ref. ²²), *NG2-DsRED* (ref. ¹⁹; JAX, no. 008241), *Aii4* (ref. ³²; JAX, no. 007914), *Ai39* (ref. ³³; JAX, no. 014539), *Ai75* (ref. ³⁴; JAX, no. 014539), *ROSA26-LSL-Mfsd2a* (generated during this study), *Thy1-GCaMP6s*³⁵ (JAX, no. 024275), *Cav1*^{-/-13} (JAX, no. 007083), *Cav1*-floxed²¹, *Nos3*^{-/-36} (JAX, no. 002684), *ROSA26-PhiC31* (ref. ³⁷; JAX, no. 007743). All mice were maintained on a mixed background and both males and females were used. For adult mice expressing *creER*, tamoxifen (Sigma-Aldrich, T5648) was dissolved in corn oil at a concentration of 20 mg ml⁻¹ and injected into peritoneal cavities with 0.2 mg g⁻¹ body weight. Six- to seven-week-old mice were treated with tamoxifen for five consecutive days and were allowed to recover for one week following the last tamoxifen treatment before cranial surgery or dissections were performed. Randomization was determined by mouse genetics as wild types, mutants and transgenic mice were assigned randomly into their respective genotype group. Sample sizes were determined by a power calculation on the basis of previous pilot data and representative sample sizes from previous literature that had similar experiments. In experiments involving mutant and transgenic mice, the genotypes were blinded until after data acquisition and analysis.

Generation of *ROSA26-LSL-Mfsd2a* transgenic mice

The targeting vector contains a CAG promoter and loxP-3×SV40PA-loxP followed by mouse *Mfsd2a* cDNA and WPRE-PolyA. A positive selection cassette, attB-PGKNeoR-attP, is located between the insertion and the 3' homologous arm which is 4.3 kb. The length of the 5' homologous arm is 1.1 kb (Extended Data Fig. 10).

The targeting vector was electroporated into embryonic stem (ES) cells derived from F₁ hybrid blastocyst of 129S6 × C57BL/6J. The G418-resistant ES clones were screened by nested PCR using primers outside the construct paired with primers inside the insertion cassette. The positive ES cell clones were used to generate chimeric mice by aggregating with 8-cell embryos of CD-1 strain. The attB-Neo-attP cassette was removed in mice by crossing the chimaeras with R26PhiC31 females (JAX, no. 007743) backcrossed in C57BL/6J for 13 generations. The F₁ pups were genotyped by PCR using primers set (5'-CCAAAGTCGCTCTGAGTTGT-3'); (5'-CCAGTTAGCCTTTAAGCCT-3') and (5'-CGGGCCATTACCGTAAGTT-3'). The PCR products are 250 bp for the wild-type allele and 329 bp for the mutant allele.

Long-term cranial window surgery

Six-week- to four-month-old mice underwent a craniotomy, implantation of a sterile glass window (3.0 mm) and attachment of a customized titanium head plate to the skull using dental cement (Metabond Parkell). Prior to the craniotomy, an intramuscular dose of dexamethasone (120 mg kg⁻¹) was administered. Mice were anaesthetized with 3–5% isoflurane and maintained at 1–2% isoflurane for the duration of the craniotomy. The respiration rate and body temperature were continuously monitored throughout the procedure to ensure the appropriate level of anaesthesia. A subcutaneous dose of the analgesic: buprenorphine (0.1 mg kg⁻¹) and ketoprofen (5.0 mg kg⁻¹) was administered at the onset of the procedure and was also administered daily for two additional days after the craniotomy. A single dose of the local anaesthetic lidocaine (20 mg kg⁻¹)/bupivacaine (2.5 mg kg⁻¹) was administered subcutaneously at the site of the craniotomy. The centre of the craniotomy over the barrel cortex and retrosplenial cortex was determined for each mouse in relation to the skull indentations bregma and lambda. Generally, for barrel cortex surgery, 3.5–3.8 mm posterior and 1.5 mm laterally from the midpoint between bregma and lambda

along the sagittal suture was marked as the centre of the craniotomy. For retrosplenial cortex surgery, the centre of the craniotomy was 3.2–3.5 mm posterior and 1.0 mm lateral from bregma. Following the craniotomy and the window with head plate implantation, mice were treated with buprenorphine/ketoprofen and observed for signs of pain and/or infection for 72 h. Furthermore, mice were handled by the experimenter, habituated to head restraint and trained to run on a foam ball daily for three consecutive days. Sensory-evoked arteriolar dilation and capillary blood flow were imaged through a 3.0-mm-diameter cranial window positioned over the somatosensory cortex in head-restrained mice with the freedom to walk on a bidirectional styrofoam ball.

Two-photon microscopy

Two-photon imaging was performed using a custom-built microscope equipped with a tunable Ti:sapphire laser (MaiTai HP DS, Spectra-Physics) controlled by ScanImage 5.1 (Vidrio Technologies). The intensity of the femtosecond pulsed infrared beam was controlled by an electro-optical modulator (Conoptics) and passed through a pair of scan mirrors (Cambridge Technology) that enabled image acquisition at 30 Hz for a field of view of 1.0 mm² and 512 × 512 pixels. Control of image zoom was enabled by controlling the resonant scanner amplitude. The objective lens used was a 16×, 0.8 NA, water-immersion lens (Nikon). Green and red fluorescence photons were separated using a custom-sized dichroic beamsplitter (580 BrightLine, Semrock) and two custom-sized single-band bandpass filters (525/50 nm BrightLine, 641/75 nm BrightLine, Semrock). Fluorescence photons were collected using photomultiplier tubes (Hamamatsu).

In vivo imaging of pial arteriolar dilation and analysis

Arterioles labelled with Alexa Fluor Hydrazide 633 were imaged at 800 nm with a field of view size of 200 μm × 200 μm (512 × 512 pixels, pixel size of 0.16 μm² per pixel) at 30 Hz. Whisker stimulation (4 Hz, 5 s) was performed using a foam brush controlled by a servo motor under the control of WaveSurfer. Alexa Fluor 633 Hydrazide (5 mg kg⁻¹) was intravenously injected into mice to visualize arterioles in vivo⁵. We imaged surface-level pial arteries and arterioles from the middle cerebral artery. Our selection of arteries and arterioles were guided by hydrazide⁺ vessels, which labels arteries and arterioles only^{5,7}. Because we are stimulating the entire whisker pad (Extended Fig. 1) as opposed to stimulating individual whiskers, we observed changes in all sampled branches. This is consistent with previous findings¹⁶ that branch orders were not necessarily relevant as long as the vessels were arterial (which they defined as SMA⁺ or hydrazide⁺). Three technical trials were acquired and averaged for each field of view. Ten to thirteen fields of view were acquired per imaging session. Three imaging sessions were collected on three separate days per mouse and arteriolar dilation responses were averaged across all three sessions for each mouse. To determine per cent change in diameter relative to baseline, the time series were first filtered with a Gaussian blur and background subtracted with a rolling ball of 50 pixels. Five-line scans orthogonal to the arterioles were sampled to generate kymographs. The two maximum intensity peaks (which represent the walls of the arterioles) were identified across the kymograph. The change in diameter of the arterioles was determined as (diameter_{time} – diameter_{baseline})/diameter_{baseline}. Diameter_{baseline} was determined as the mean diameter during the 3 s before the whisker stimulation. Diameter_{time} is the vessel diameter at a particular time. The change in maximum diameter was determined as the maximum value during the whisker stimulation. To determine latency onset to dilate, a line was fitted through 80% and 20% of the maximum value. The latency onset to dilate was considered the time difference between the x-intercept of the line and the start of the whisker stimulation.

In vivo imaging of parenchymal arteriolar dilation and analysis

Arterioles stained with Alexa Fluor 633 Hydrazide (5 mg kg⁻¹) were imaged at multiple depths within the barrel cortex. Three parenchymal

Article

depths were used at 100 μm , 200 μm and 300 μm from the pial surface. Three technical trials were acquired and averaged for each field of view at each depth location. Three-to-four diving arterioles were imaged per imaging session and three sessions on sequential days were recorded in total per mouse. For analysis of parenchymal arterial dilation which appear as ellipses, we tracked the diving arterioles by fitting an ellipse to the hydrazide signal. Movies were first averaged using a three-frame rolling average then smoothed with a Laplacian of Gaussian filter. Intensities of the smoothed movie were rescaled to range from 0 to 255 and a threshold pixel intensity was picked to exclude background fluorescence signal outside of the ring of hydrazide signal surrounding the arteriole. Next, ridge detection was performed using the FIJI ridge detection plugin³⁸. Three separate ridge-detection parameter sets were used and the ridges were combined to detect bright and dim ridges. Next, an initial estimate of the ellipse fit was determined using an elliptical Hough transform on the binarized ridge image. The fit was then refined by minimizing the distance between the ridge pixels and the fitted ellipse using the Hough transform results as the starting parameter values. The ellipse was parameterized as follows:

$$x(\alpha) = a \cos(\alpha)\cos(\theta) - b \sin(\alpha)\sin(\theta) + x_0$$

$$y(\alpha) = a \cos(\alpha)\sin(\theta) - b \sin(\alpha)\cos(\theta) + y_0$$

in which $x(\alpha)$, $y(\alpha)$ are the x and y coordinates of the points on the ellipse, a and b are the two ellipse axes, θ is the tilt angle of the ellipse, x_0 and y_0 are the coordinates of the ellipse centre, and α ranges from 0 to 2π to circumscribe the entire ellipse perimeter. Minimization was done in MATLAB using the lsqnonlin function and 95% confidence intervals of the parameters were determined using nlparci. To be consistent with the pial artery dilation tracking, we report changes in diameter of the parenchymal arterioles ($\Delta D/D_{\text{baseline}}$) from the minor axis of the ellipse, which does not depend on the orientation of the arteriole cross section relative to the microscope optical axis. As image quality is variable, frames with poor accuracy fitting are discarded by first rejecting frames in which the minor axis fit is more than two median absolute deviations from the ten-frame sliding window median. Next, frames that have a confidence interval greater than four pixels for the minor axis are discarded. If fewer than 50% of the frames of the trajectory remain, the entire trajectory is discarded. Finally, the trajectory was smoothed using the MATLAB smooth function with rlwess. To obtain $\Delta D/D_{\text{baseline}}$, this value was multiplied by two and divided by the mean diameter immediately 3 s before whisker stimulation. The three technical replicates were averaged as with the pial artery dilation experiments to produce the trajectory for the vessel.

In vivo imaging of capillary red blood cell velocity and analysis

Mice were intravenously injected with quantum dots 525 (Thermo Fisher Scientific) and Alexa Fluor Hydrazide 633 to unambiguously distinguish arterioles and capillaries in vivo. Hydrazide-negative capillaries were imaged with a field of view size of 100 $\mu\text{m} \times 25 \mu\text{m}$ (512 \times 25 pixels, pixel size of 0.04 μm^2 per pixel) at 610 Hz. Whisker stimulation (4 Hz, 5 s) was performed using a foam brush controlled by a servo motor under the control of Wavesurfer. Three technical trials were acquired and averaged for each field of view. Three imaging sessions were collected on three separate days per mouse and changes in capillary red blood cell velocity were averaged across all three sessions for each mouse. To determine per cent change in velocity relative to baseline, the movies were first filtered with a Gaussian blur and background subtracted with a rolling ball of 50 pixels. Five-line scans parallel to the flow of red blood cells were sampled to generate kymographs. Using a published algorithm³⁹ that uses an iterative radon transform and edge detection filter, the change in velocity of red blood cells was determined as $(\text{velocity}_{\text{time}} - \text{velocity}_{\text{baseline}})/\text{velocity}_{\text{baseline}}$. Velocity baseline was defined as the mean velocity during the 3 s before the whisker

stimulation. $\text{Velocity}_{\text{time}}$ is the velocity of the red blood cell flow at that moment in time. The change in maximum velocity was determined as the maximum value during the whisker stimulation. To determine latency onset to increase red blood cell velocity, a line was fitted through 80% and 20% of the maximum velocity value. The latency onset was considered the time difference between the x intercept of the fitted line and the start time of the whisker stimulation.

In vivo two-photon imaging and pharmacology

To assess vasodilatory function of pial artery SMCs, we imaged pial artery diameter changes with two-photon microscopy in response to topical application of the nitric oxide donor DEA NONOate (EMD Millipore). Arteries were stained with Alexa Fluor 633 Hydrazide, as described earlier, one day before the imaging session. Under isoflurane anaesthesia (1.0–1.5%), a titanium head-plate with a 10-mm-diameter hole centred over the right parietal skull bone was cemented onto the skull using Metabond. A custom perfusion system was constructed along the rim of the head-plate hole to allow for simultaneous imaging and application of DEA NONOate. A 5.0–7.0-mm circular craniotomy of the right parietal skull bone was carefully performed and the newly exposed cortex was kept submerged in artificial cerebral spinal fluid (aCSF) for the duration of the experiment. The mouse was then carefully transitioned from isoflurane (0.5%) to ketamine/xylazine (100 mg kg^{-1}) and transferred from the surgical stereotaxic stage to the imaging platform where body temperature was maintained at 37 $^{\circ}\text{C}$ using a heat pad. A fresh 1.0 μM DEA NONOate solution was prepared immediately before the start of the imaging session given the short half-life of DEA NONOate in aqueous solution. A syringe pump (Harvard Apparatus) controlled the application of the DEA NONOate solution onto the exposed cortex at 1.0 ml min^{-1} . Five recordings (30 s in duration) of each pial vessel per mouse were collected in series interleaved with 30 s of washing with aCSF. The five measurements were then averaged per mouse.

Ex vivo, acute slice two-photon slice imaging and pharmacology

Acute coronal brain slices were prepared by deeply anaesthetizing mice with isoflurane inhalation followed cardiac perfusion with ice-cold choline-based cut solution containing (in mM): 25 NaHCO_3 , 25 glucose, 1.25 NaH_2PO_4 , 7 MgCl_2 , 2.5 KCl , 0.5 CaCl_2 , 11.6 ascorbic acid, 3.1 pyruvic acid and 110 choline chloride. After brain dissection and blocking, 300- μm slices were prepared in cut solution with a Leica VT1000 s vibratome. Slices were then transferred for 30 min to recovery into a holding chamber containing 34 $^{\circ}\text{C}$ (aCSF) containing (in mM): 125 NaCl , 2.5 KCl , 1.25 NaH_2PO_4 , 25 NaHCO_3 , 11 glucose, 2 CaCl_2 and 1 MgCl_2 . During recovery, slices were incubated with approximately 1 μM Alexa Fluor 633 Hydrazide (ThermoFisher). Following recovery, slices were imaged while constantly perfused with room temperature aCSF. Choline cut solution and aCSF were constantly bubbled with 5% CO_2 /95% O_2 . Imaging was performed on a custom-built two-photon microscope and images acquired with a custom version of ScanImage written in MATLAB (Mathworks). During imaging, arteries were constricted with 100 nM U46619 (Sigma-Aldrich) and dilated by acutely dissolving about 5 mg of DEA NONOate (EMD Millipore) into the 10 ml of recycling aCSF being perfused over the slice.

Non-invasive blood pressure measurement procedure in awake mice

Systolic, diastolic and mean blood pressure were measured using the non-invasive tail-cuff method (CODA Monitor, Kent Scientific). At the start of a blood pressure measurement session, the appropriate mouse holder was selected on the basis of the mouse's weight. The holder was placed over a heating pad with a set temperature that is regulated at 38 $^{\circ}\text{C}$. The tail-cuff and volume pressure recording sensor were placed on the heating pad and covered with a blanket to allow these components to reach the set temperature. After 2–3 min, the awake mouse was gently introduced into the holder. A light blanket was draped over

the tail and the mouse was left alone for 3–5 min to allow habituation. The blood pressure measurements take place via 10–20 tail-cuff inflation–deflation sweeps that in total take from 5–10 min in duration. Multiple days of measurements may be required to gain confidence in the accuracy of the measurements. After the measurements, the mouse was carefully removed from the holder and immediately placed in its cage. Measurements of all sweeps are then averaged per mouse.

SMC coverage quantification

NG2^{DsRED} (or *CSPG4^{DsRED}*) mice were crossed to *Cav1^{-/-}* mice. Cranial window surgeries were performed over the barrel cortex of *Cav1^{+/+};NG2^{DsRED+}* and *Cav1^{-/-};NG2^{DsRED+}* mice. Mice were injected with Alexa Fluor 488 Hydrazide (5 mg kg⁻¹). Around 30 arterioles (hydrazide⁺DsRed⁺) per mouse were imaged. A 100-μm intensity line profile was drawn perpendicularly to the contractile bands of the SMCs. Maximum peaks corresponding to individual smooth muscle were counted and the number of SMCs per 100 μm length was determined.

Transmission electron microscopy

Brains from adult mice were dissected and fixed by immersion in 5% glutaraldehyde, 4% PFA and 0.1 M sodium cacodylate for 2 weeks at room temperature. Following fixation, brains were washed overnight in 0.1 M sodium cacodylate. Coronal vibratome free-floating sections of 50 μm were collected. The cortices, particularly somatosensory and motor, were microdissected, post-fixed in 1% osmium tetroxide and 1.5% potassium ferrocyanide, dehydrated, and embedded in epoxy resin. Ultrathin sections of 80 nm were then cut from the block surface, collected on copper grids, and counter-stained with Reynold's lead citrate and examined under a 1200EX electron microscope (JEOL) equipped with a 2k CCD digital camera (AMT).

Mean vesicular density

For all transmission electron microscopy quantifications, mean vesicular density values were calculated from the number of vesicles per μm² of cell area for each image collected. All images were collected at 12,000× magnification and analysis was performed blinded. Each density value (circle on the graphs) represents an individual vessel (capillary or arteriole). The same colour of the circle represents vessels analysed from the same mouse. Values are expressed as mean ± s.e.m.

Immunohistochemistry

Mice were anaesthetized with ketamine/xylazine via intraperitoneal injection and then mice were transcardially perfused with cold PBS and followed by cold 4% PFA. Brains and retinas were fixed by immersion in 4% PFA/PBS overnight at 4 °C. Next, brains and retinas were washed 3× in PBS. Brain sections were either cut as 50-μm sections on the vibratome or cryopreserved in 30% sucrose, frozen in TissueTek OCT (Sakura) and cut as 25-μm sections on the cryostat. For MFSD2A and eNOS immunohistochemistry, mice were euthanized by cervical dislocation. Brains were snap-frozen with liquid nitrogen and cut as 25-μm sections on the cryostat. Brain sections were fixed with chilled methanol for 10 min. Brain sections and retinas were blocked with 10% goat or donkey serum, 5% BSA, PBST (0.5% Triton X-100) and stained overnight at 4 °C with the following primary antibodies at the indicated concentrations: MFSD2A (1:200, Cell Signaling Technologies; RRID: AB_2617168) or MFSD2A (a gift from D. Silver, as used previously⁴⁰), SMA (1:1,000, Sigma-Aldrich, no. C6198, RRID: AB_476856), ICAM2 (1:200, BD Biosciences, no. 553326, RRID: AB_394784), PECAM (1:200, BD Biosciences, no. 553370; RRID: AB_394816), claudin 5 (Thermo Fisher Scientific, no. 34-1600, RRID: AB_2533157), eNOS (Abcam, no. ab5589; RRID: AB_304967), MYH11 (Abcam, no. ab53219; RRID: AB_2147146) or desmin (ThermoFisher, no. PA1-37556) and TAGLN (Abcam, no. ab14106), followed by corresponding Alexa Fluor-conjugated secondary antibodies (1:500, ThermoFisher) and Alexa Fluor Hydrazide

633 (1:1,000, ThermoFisher). Tissues were mounted with ProLong Gold for imaging.

In situ NO detection

This assay was previously adapted from these studies^{41,42}. Mice were deeply anaesthetized with ketamine/xylazine. Mice were transcardially perfused with warm (37 °C) 50 ml PBS, then perfused with 50 ml warm PBS containing 10 μM DAF-2 (Thermo Fisher Scientific, D23842), 100 μM L-arginine and 2 mM CaCl₂. Next, mice were perfused with warm PBS again followed by 4% PFA. Brains and kidneys were collected and fixed overnight in 4% PFA at 4 °C. Tissues were sectioned on the vibratome (50 μm) and processed for immunostaining.

Light microscopy

Olympus FluoView FV1200 and Leica SP8 laser scanning confocal microscopes (20×, 0.75 NA, 40×, 1.3 NA, 63× 1.4 NA) and an Olympus VS 120 slide scanner (10×, 0.4 NA) were used for imaging retina flat-mounts and brain sections. Images were processed using Adobe Photoshop, Illustrator, Olympus Fluoview and FIJI (NIH).

Statistical analyses

All statistical analyses were performed using Prism 7 and 8 (GraphPad Software). Two group comparisons were analysed using an unpaired two-tailed Student's *t*-test or non-parametric analyses. Multiple group comparisons were analysed using a one-way ANOVA, followed by a post hoc Bonferroni analysis to correct for multiple comparisons. No data were excluded when performing statistical analysis. The s.e.m. was calculated for all experiments and displayed as errors bars in graphs. Statistical details for specific experiments—including exact *n* values and what *n* represents, precision measures, statistical tests used and definitions of significance—can be found in figure legends. Each colour circle on the graphs throughout the study represents an individual vessel (capillary or arteriole or SMC). The same colour of the circle represents vessels analysed from the same mouse. Values are expressed as mean ± s.e.m. Please see Supplementary Table 1 for statistical test results.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source Data for quantification described in the text or shown in graphs plotted in Figs. 1–4 and Extended Data Fig. 1–10 are available with the paper.

Code availability

The source code to run the pial arteriolar dilation analysis is available at <https://github.com/gulabneuro/Pial-Vasodilation-Analysis>. The source code to run the parenchymal arterial dilation analysis is available at <https://github.com/gulabneuro/divingArterioleTracking>.

31. Pu, W. et al. Mfsd2a⁺ hepatocytes repopulate the liver during injury and regeneration. *Nat. Commun.* **7**, 13369 (2016).
32. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
33. Madisen, L. et al. A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat. Neurosci.* **15**, 793–802 (2012).
34. Quina, L. A., Harris, J., Zeng, H. & Turner, E. E. Specific connections of the interpeduncular subnuclei reveal distinct components of the habenulopeduncular pathway. *J. Comp. Neurol.* **525**, 2632–2656 (2017).
35. Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
36. Shesely, E. G. et al. Elevated blood pressures in mice lacking endothelial nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **93**, 13176–13181 (1996).

37. Raymond, C. S. & Soriano, P. High-efficiency FLP and ΦC31 site-specific recombination in mammalian cells. *PLoS ONE* **2**, e162 (2007).
38. Steger, C. An unbiased detector of curvilinear structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 113–125 (1998).
39. Chhatbar, P. Y. & Kara, P. Improved blood velocity measurements with a hybrid image filtering and iterative Radon transform algorithm. *Front. Neurosci.* **7**, 106 (2013).
40. Nguyen, L. N. et al. Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid. *Nature* **509**, 503–506 (2014).
41. Kanetsuna, Y. et al. Deficiency of endothelial nitric-oxide synthase confers susceptibility to diabetic nephropathy in nephropathy-resistant inbred mice. *Am. J. Pathol.* **170**, 1473–1484 (2007).
42. Jiang, R. et al. Generation of a conditional allele for the mouse endothelial nitric oxide synthase gene. *Genesis* **50**, 685–692 (2012).

Acknowledgements We thank C. Harvey for his help in the design and construction of the two-photon microscope; O. Mazar and P. Gorelik at the HMS Research Instrumentation Core Facility and T. LaFratta and J. LeBlanc of the Harvard Neuroengineering and Imaging Machine Shop for their help in the construction of the two-photon microscope; S. Ashrafi, J. Cohen, D. Ginty, C. Weitz., G. Yellen and members of the Gu laboratory for comments on the manuscript; C. Lahmann for protocol advice and assistance on cranial window surgery; P. Scherer, R. Adams and B. Zhou for the Cav1-floxed, *BMX^{creER}* and *Mfsd2a^{creER}* mice, respectively; D. Silver for the

MFS2A antibody; and the HMS Electron Microscopy Core Facility, HMS Neurobiology Imaging Facility and HMS NeuroDiscovery Center for consultation and instrument availability. This work was supported by Quan Fellowship (B.W.C.), NIH T32 and the Mahoney Postdoctoral Fellowship (V.N.), Jane Coffin Childs Fund (A.J.G.), K99 NS102429 (A.J.G.), R37 NS046579 (B.L.S.), P30NS072030 (HMS Neurobiology Imaging Facility), the NIH DP1 NS092473 Pioneer Award (C.G.), Fidelity Biosciences Research Initiative (C.G.). The research of C.G. was also supported in part by a Faculty Scholar grant from the Howard Hughes Medical Institute.

Author Contributions B.W.C. and C.G. conceived the project. B.W.C., V.N. and C.G. designed experiments. B.W.C., V.N., A.J.G., K.B. and H.L.Z. performed experiments. B.W.C., V.N., A.J.G., L.K. and P.K. analysed all data. B.W.C. and C.G. wrote the manuscript, with feedback from all authors.

Competing interests The authors declare no competing interests.

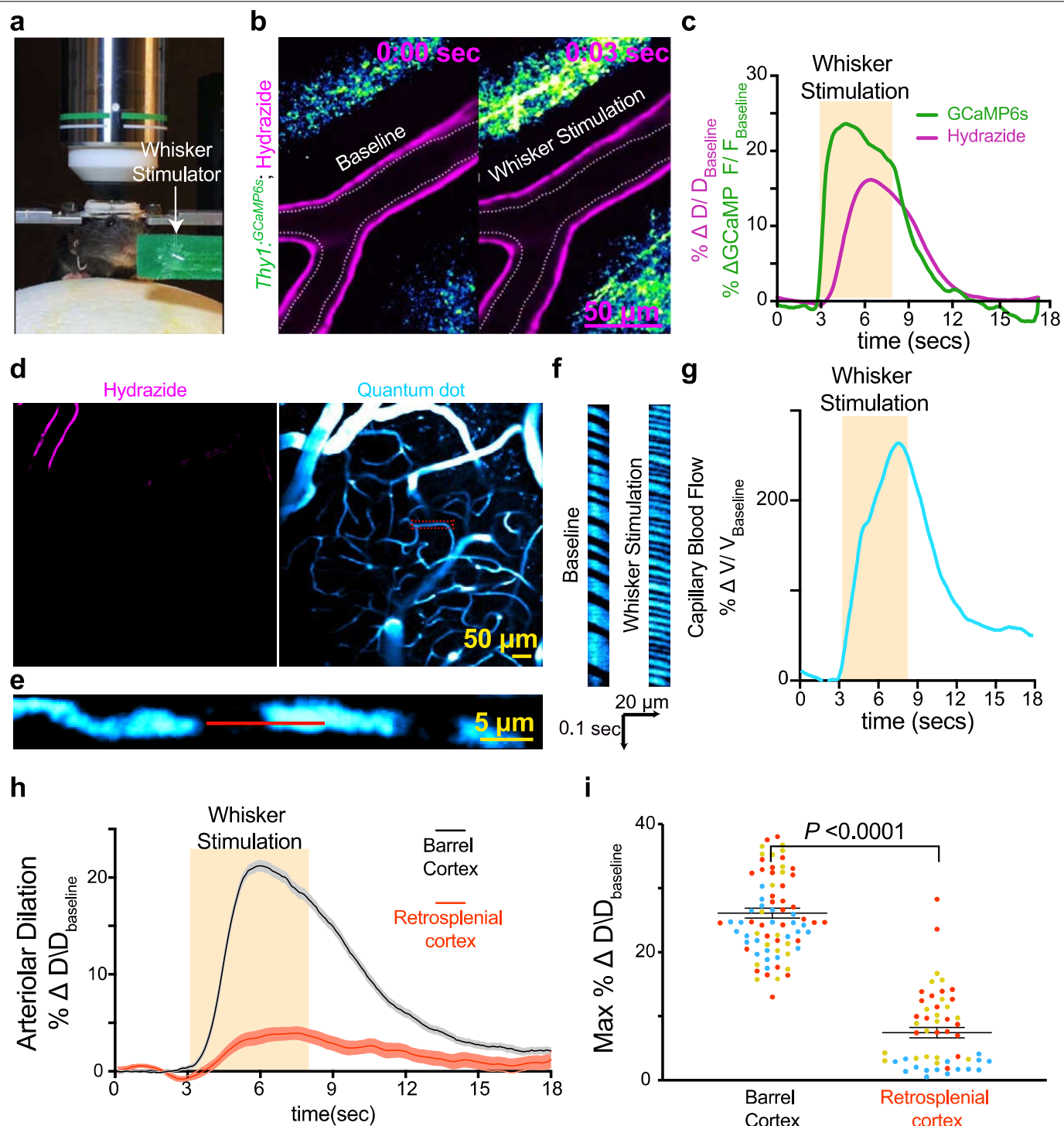
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2026-1>.

Correspondence and requests for materials should be addressed to C.G.

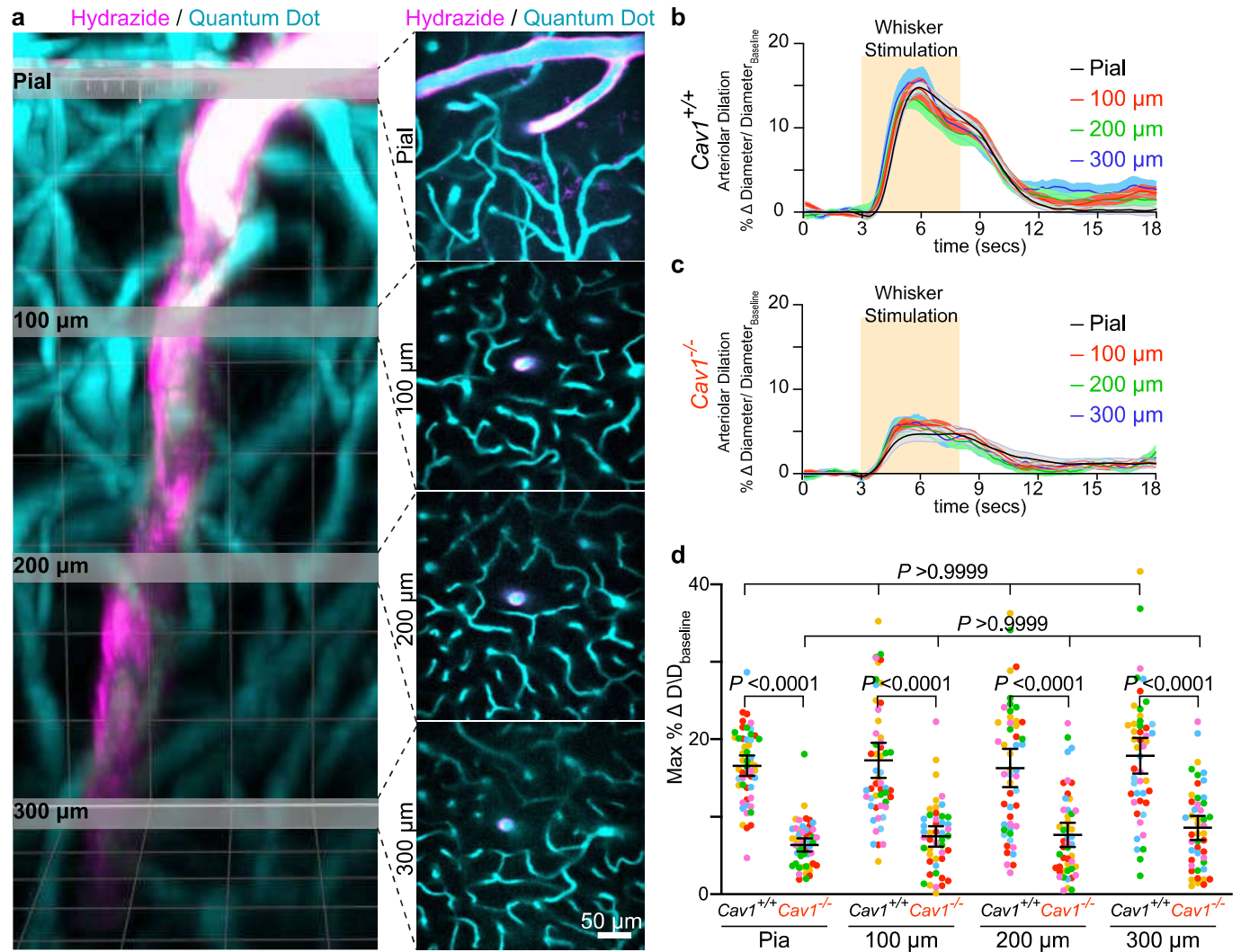
Peer review information *Nature* thanks Dritan Agalliu, Brian MacVicar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



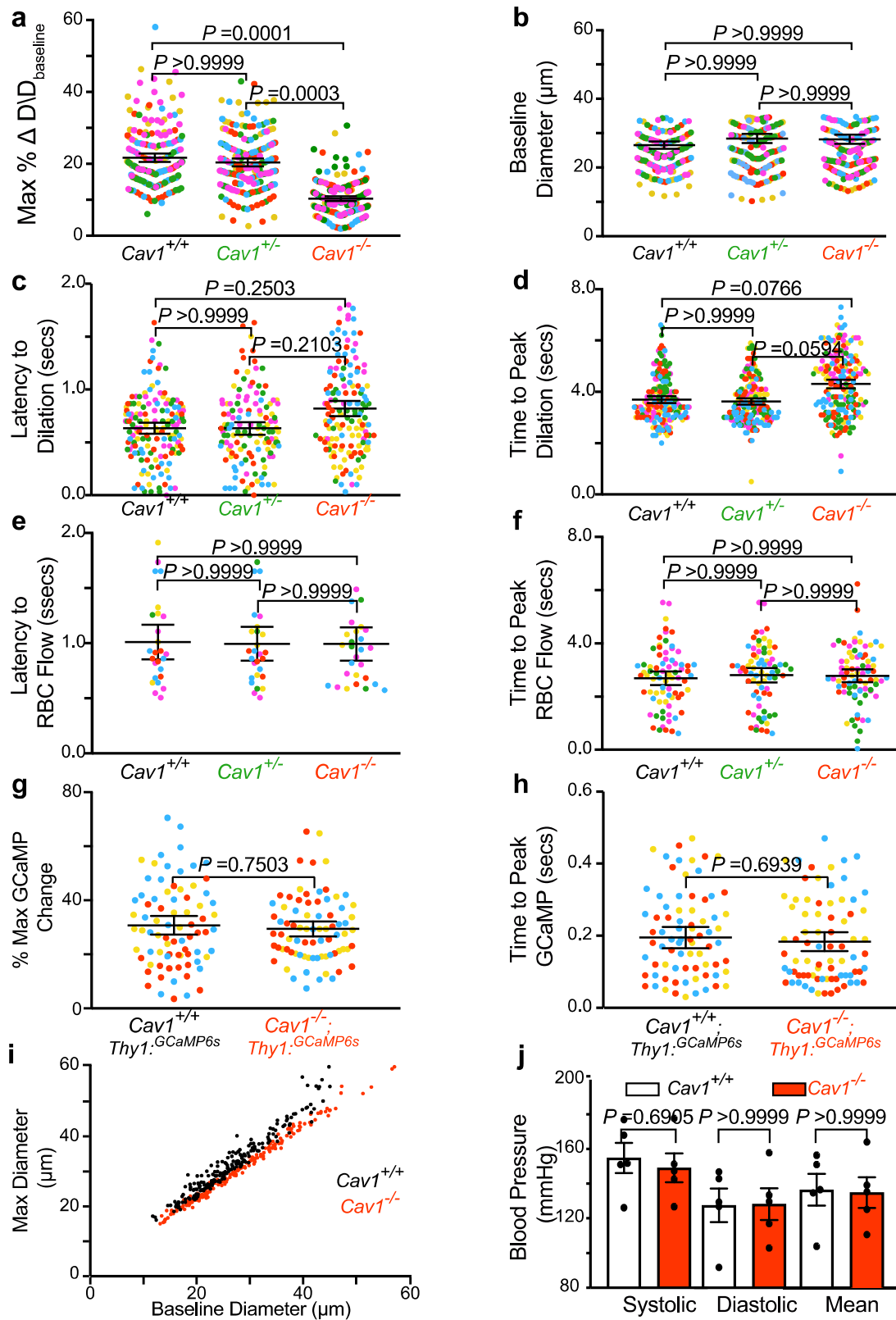
Extended Data Fig. 1 | In vivo two-photon imaging of neurovascular coupling in the barrel cortex and retrosplenial cortex. **a**, Setup of the in vivo microscopy. Awake mice with cranial windows over the barrel cortex are head-fixed and allowed to move on a foam ball. Whisker stimulator (arrow) is used for brushing whiskers to evoke neural activity in the barrel cortex. **b–g**, Imaging in the barrel cortex. **b**, Hydrazide injection in *Thy1-GCaMP6s* mice enables simultaneous imaging of neural activity (green) and arteriolar dilation (magenta). Two-photon imaging of arterioles and neural activity before (left) and after (right) whisker stimulation. Hashes indicate the baseline diameter at time = 0 s. **c**, Time course of change in arteriolar dilation (magenta) and GCaMP6s fluorescence (green). Orange bar signifies the period of whisker stimulation. **d**, Two-photon imaging of arterioles (magenta) and capillary blood flow (blue). After intravenous injection of quantum dots, the plasma is

bright whereas the red blood cells are dark. **e**, High magnification of a capillary outlined by the red box in **d**. Minimizing the image size increases the temporal resolution to about 610 Hz or 1.6 ms per frame. **f**, Kymographs of capillary blood flow during baseline (left) and whisker stimulation (right). Kymographs were generated from the parallel line scan (red line) of the capillary blood flow in **e**. **g**, Time course of change in red blood cell velocity. **h**, Time course of change in arteriolar dilation in the barrel cortex (black, $n = 78$ arterioles, 3 mice) and in the retrosplenial cortex (red, $n = 54$ arterioles, 3 mice). **i**, Maximum percentage change in arteriolar dilation upon whisker stimulus in these two brain regions upon whisker stimulus. The orange bar signifies the period of whisker stimulation. Data are mean \pm s.e.m.; nested unpaired, two-tailed t -test for **i**.



Extended Data Fig. 2 | *Cav1*-knockout mice have impaired vasodilation in both pial arteries and penetrating arterioles diving deep into the parenchyma. **a**, Three-dimensional volume rendering of a two-photon-imaged site in *Cav1*^{+/+} mouse barrel cortex, from the pial surface to a depth of about 400 μm . The lumen of all vessels is filled with quantum dots (blue) and arterioles are labelled with hydrazide (magenta). The deepest imaged bin is at 300 μm because we see the appearance of the hydrazide start at 300 μm , indicating that this is at the start of the arteriolar vessels. This observation is also consistent with a previous study⁵, which characterized hydrazide as an arteriolar vessel marker. Grey slices correspond to z cross-sections shown per depth. Independent replicates for **a** were performed in five wild-type mice.

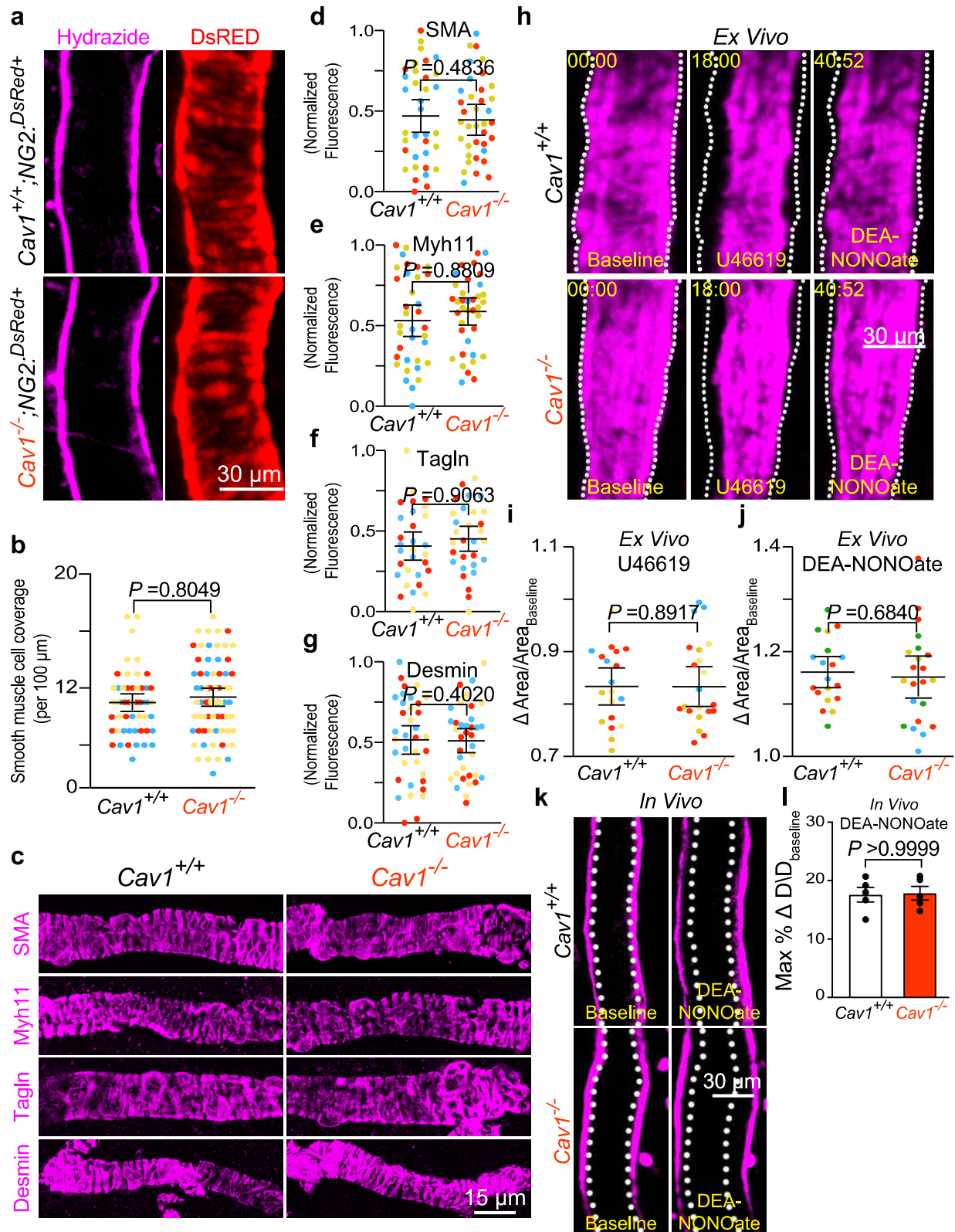
b, c, Time course of change in arteriolar dilation in the barrel cortex from *Cav1*^{+/+} ($n = 5$ mice, 10–15 arterioles per depth) (**b**) and *Cav1*^{-/-} mice ($n = 5$ mice, 10–15 arterioles per depth) (**c**). **d**, Maximum percentage change in arteriolar dilation upon whisker stimulation between *Cav1*^{+/+} and *Cav1*^{-/-} mice at the indicated depth. Statistical significance was determined by two-way ANOVA with a post hoc Bonferroni multiple comparison adjustment for **d**. All data are mean \pm s.e.m. We compared the maximum percentage change in arteriolar dilation upon whisker stimulation between *Cav1*^{+/+} and *Cav1*^{-/-} mice at each depth and also compared the responses across depth within the same genotype.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | *Cav1*-knockout mice have attenuated vasodilation but normal neural activity and neurovascular coupling kinetics. **a–d**, Maximum percentage change in dilation response (**a**) and baseline diameter (**b**) latency to maximum change in arteriolar dilation (**c**), time to peak dilation (**d**) in *Cav1*^{+/+} (*n* = 193 arterioles, 40 capillaries, 5 mice), *Cav1*^{+/-} (*n* = 123 arterioles, 40 capillaries, 5 mice), and *Cav1*^{-/-} mice (*n* = 153 arterioles, 31 capillaries, 5 mice). **e, f**, Latency to maximum red blood cell flow velocity (**e**) and time to peak red blood cell flow (**f**) in *Cav1*^{+/+} (*n* = 193 arterioles, 40 capillaries, 5 mice), *Cav1*^{+/-} (*n* = 123 arterioles, 40 capillaries, 5 mice) and *Cav1*^{-/-} mice (*n* = 153 arterioles, 31 capillaries, 5 mice). **g, h**, Maximum percentage change in GCaMP6s (**g**) and

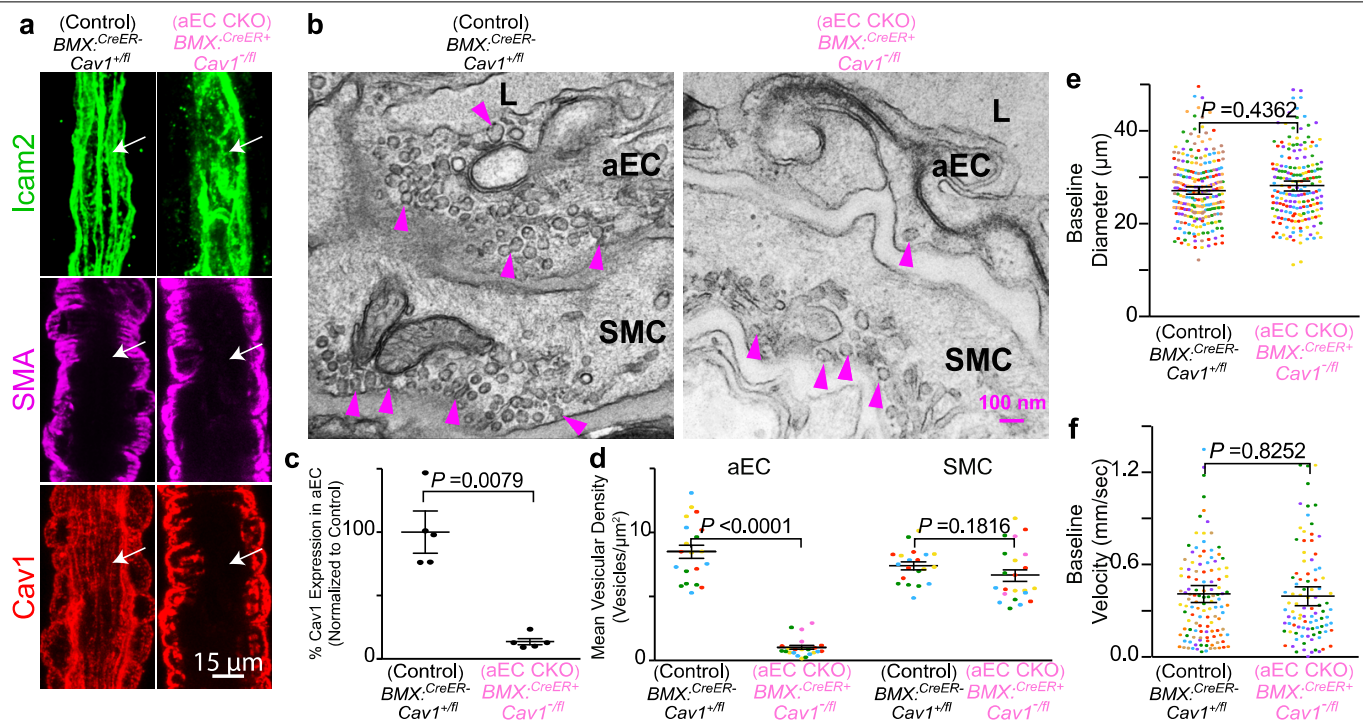
latency to peak change in GCaMPs (**h**) in *Cav1*^{+/+}; *Thy1*-GCaMP6s (*n* = 78 field of views of the neuropils, 5 mice) and *Cav1*^{-/-}; *Thy1*-GCaMP6s (*n* = 78 neuropils, 5 mice). Each circle represents an individual trial of GCaMP6s signal. **i**, Baseline diameter to absolute maximum diameter response during whisker stimulation in *Cav1*^{+/+} and *Cav1*^{-/-} mice. **j**, Tail-cuff blood pressure measurements between *Cav1*^{+/+} (*n* = 5 mice) and *Cav1*^{-/-} mice (*n* = 5 mice). Statistical significance was determined by a one-way nested ANOVA with a post hoc Bonferroni multiple comparison adjustment for **a–f**, a nested unpaired, two-tailed *t*-test for **g, h**, and two-tailed Mann–Whitney *U* test for **j**. All data are mean ± s.e.m.



Extended Data Fig. 4 | See next page for caption.

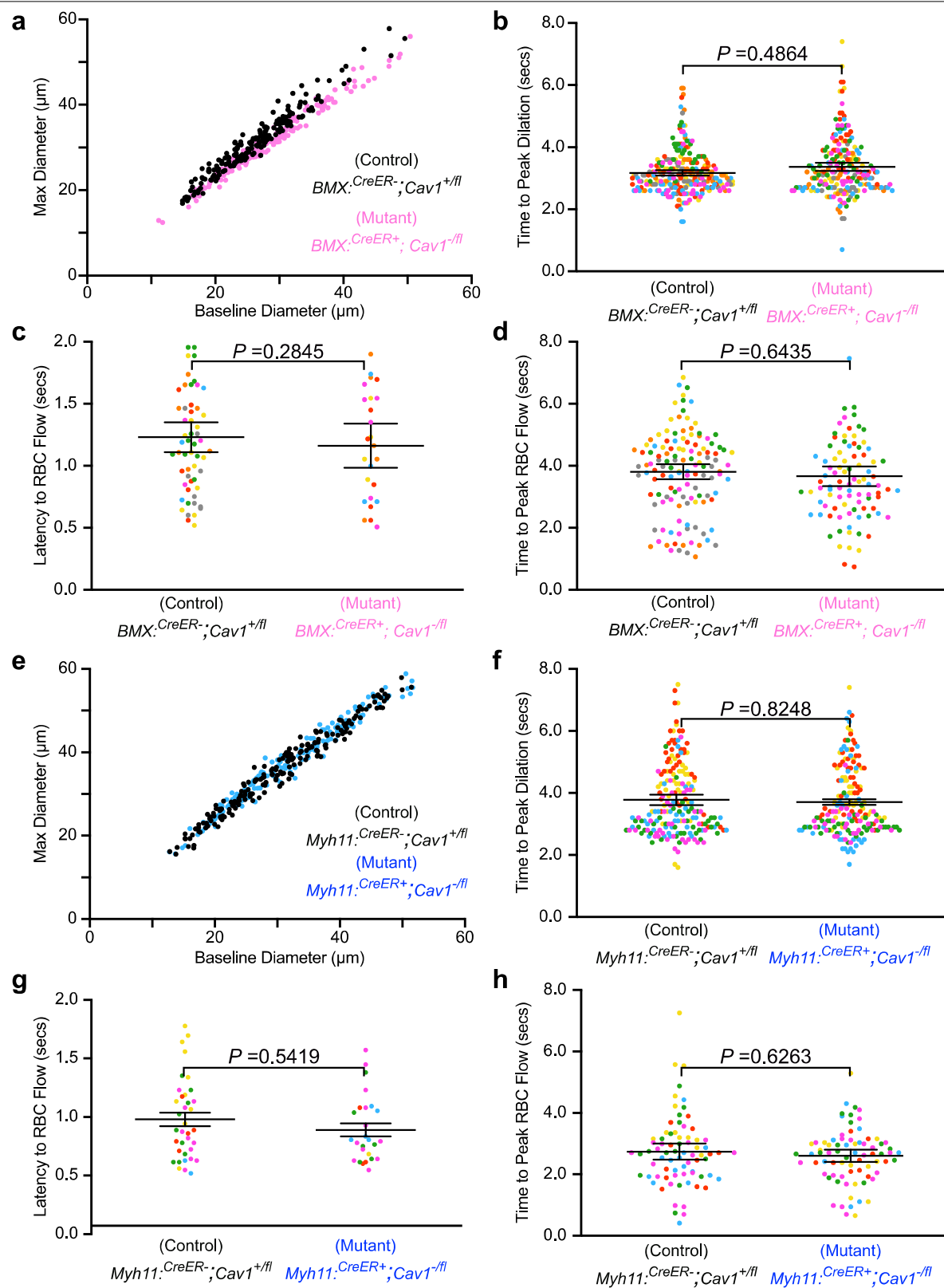
Extended Data Fig. 4 | *Cav1*-mutant mice exhibit normal SMC integrity and function. **a**, In vivo two-photon microscopy images of hydrazide (magenta) and DsRed (red) from *Cav1^{+/+}NG2^{DsRED+}* and *Cav1^{-/-}NG2^{DsRED+}* mice. **b**, Quantification of DsRED⁺ SMCs per 100 μ m as shown in **a**, in *Cav1^{+/+}* ($n = 3$ mice, 27 arterioles) and *Cav1^{-/-}* ($n = 3$ mice, 28 arterioles) mice. **c**, Immunostaining for SMC contractile proteins, including SMA, MYH11, TAGLN and desmin on brain arterioles from *Cav1^{+/+}* and *Cav1^{-/-}* mice. **d–g**, Normalized fluorescence quantification of the various contractile proteins from *Cav1^{+/+}* and *Cav1^{-/-}* mice. **h**, Still frame images of arterioles labelled with hydrazide (magenta) in ex vivo acute brain slices from *Cav1^{+/+}* and *Cav1^{-/-}* mice using two-photon microscopy. Left, arterioles during baseline; middle, arterioles during U46619 (thromboxane agonist) treatment; right,

arterioles during DEA NONOate (NO donor) treatment. White hashes outline the arterioles during baseline based on time = 0 min. **i, j**, Maximum arteriolar contraction by U46619 (**i**) and maximum arteriolar dilation by DEA NONOate (**j**) on acute brain slices from *Cav1^{+/+}* ($n = 5$ mice, 19 arterioles) and *Cav1^{-/-}* ($n = 5$ mice, 22 arterioles). **k**, In vivo images of arterioles labelled with hydrazide (magenta) from *Cav1^{+/+}* and *Cav1^{-/-}* mice using two-photon microscopy. Left, arterioles during baseline; right, arterioles during DEA NONOate superfusion. White hashes outline the arterioles during baseline based on time = 0 s. **l**, Quantification of maximum arteriolar dilation during DEA NONOate superfusion in vivo ($n = 5$ mice for both genotypes). Statistical significance was determined by nested, unpaired, two-tailed t -test for **b, d–g, i, j**, and by two-tailed Mann–Whitney U test for **l**. Data shown as mean \pm s.e.m.



Extended Data Fig. 5 | Caveolae in CNS aECs are abolished in aEC conditional *Cav1*-knockout mice. **a**, Immunostaining of adult brain sections for ECs (ICAM2, green), SMCs (SMA, magenta) and CAV1 (red) from control (*BMX^{CreER}-;Cav1^{+/fl}*) and aEC-specific conditional CAV1 mutant (*BMX^{CreER}+;Cav1^{-fl}*) mice. Arrows point to aECs. **b**, Transmission electron microscopy images of CNS aECs and SMCs from control and aEC-specific conditional *Cav1*-mutant mice. Arrowheads point to caveolae. L, Lumen. **c**, Quantification of mean normalized immunofluorescence of CAV1 in aECs from control ($n=5$ mice) and aEC-specific conditional *Cav1*-mutant mice ($n=5$ mice). **d**, Quantification of the

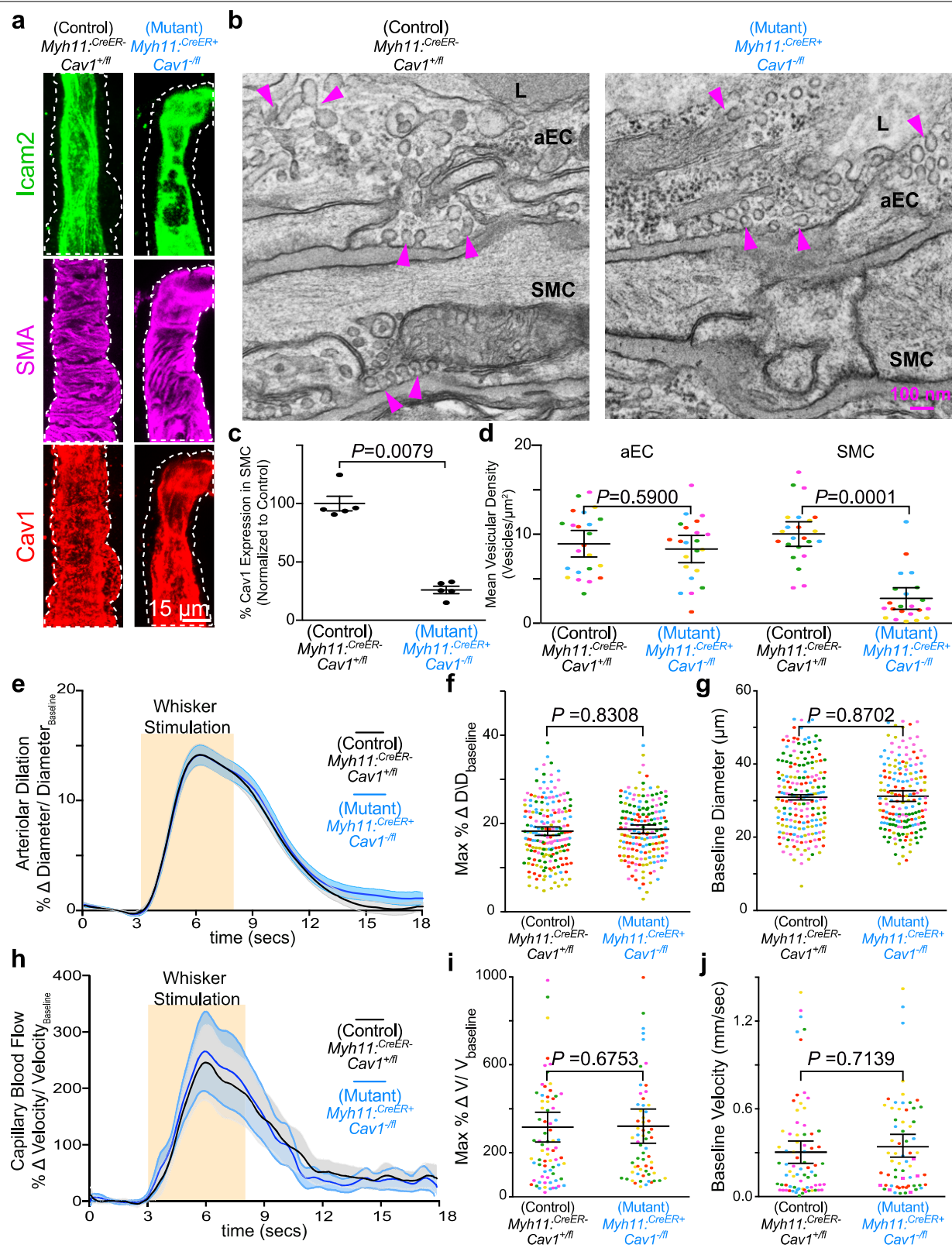
mean vesicular density in aECs and SMCs between control ($n=4$ mice, 20 arterioles) and aEC-specific conditional *Cav1*-mutant mice ($n=5$ mice, 22 arterioles). **e**, **f**, Quantification of baseline diameter (**e**) and baseline velocity in control (*BMX^{CreER}-;Cav1^{+/fl}*, $n=7$ mice, 260 arterioles, 122 capillaries) and aEC conditional *Cav1*-knockout mice (*BMX^{CreER}+;Cav1^{-fl}*, $n=5$ mice, 193 arterioles, 94 capillaries). Statistical significance was determined by Mann-Whitney test for (**c**) and nested, unpaired, two-tailed *t*-test for (**d-f**). Data are shown as mean \pm s.e.m.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Conditional aEC-specific and SMC *Cav1*-knockout mice have normal neurovascular coupling kinetics. **a**, Baseline diameter to absolute maximum diameter response during whisker stimulation in control ($BMX^{creER-/-};Cav1^{+/f1}$) and mutant ($BMX^{creER+};Cav1^{-/f1}$) mice. **b**, Quantification of time to peak arteriolar dilation in control ($BMX^{creER-/-};Cav1^{+/f1}$; $n = 7$ mice; 234 arterioles) and aEC-specific conditional *Cav1*-mutant ($BMX^{creER+};Cav1^{-/f1}$; $n = 5$ mice; 202 arterioles) mice. **c**, Quantification of latency to peak red blood cell flow velocity in control ($BMX^{creER-/-};Cav1^{+/f1}$; $n = 7$ mice; 58 capillaries) and aEC-specific conditional *Cav1*-mutant ($BMX^{creER+};Cav1^{-/f1}$; $n = 5$ mice; 25 capillaries) mice. **d**, Quantification of time to peak red blood cell flow velocity in control ($BMX^{creER-/-};Cav1^{+/f1}$; $n = 7$ mice; 127 capillaries) and aEC-specific conditional *Cav1*-mutant ($BMX^{creER+};Cav1^{-/f1}$; $n = 5$ mice; 94 capillaries) mice. **e**, Baseline diameter

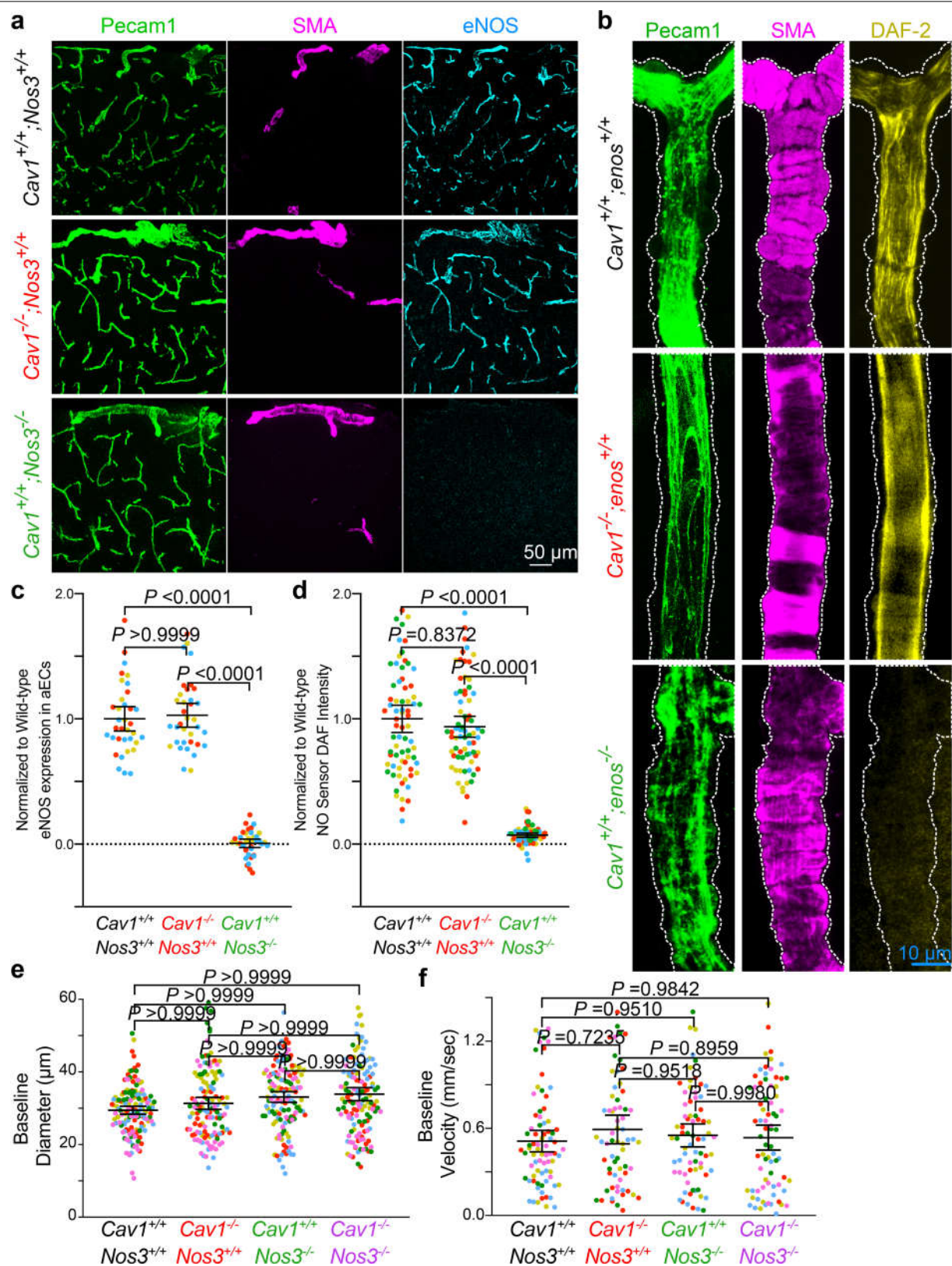
to absolute maximum diameter response during whisker stimulation in control ($Myh11^{creER-/-};Cav1^{+/f1}$) and mutant ($Myh11^{creER+};Cav1^{-/f1}$) mice. **f**, Quantification of time to peak arteriolar dilation in control ($Myh11^{creER-/-};Cav1^{+/f1}$; $n = 5$ mice; 193 arterioles) and SMC conditional *Cav1*-mutant ($Myh11^{creER+};Cav1^{-/f1}$; $n = 5$ mice; 180 arterioles) mice. **g**, Quantification of latency to red blood cell flow in control ($Myh11^{creER-/-};Cav1^{+/f1}$; $n = 5$ mice; 36 capillaries) and SMC conditional *Cav1*-mutant ($Myh11^{creER+};Cav1^{-/f1}$; $n = 5$ mice; 26 capillaries) mice. **h**, Quantification time to peak red blood cell flow velocity in ($Myh11^{creER-/-};Cav1^{+/f1}$; $n = 5$ mice; 75 capillaries) and SMC conditional *Cav1*-mutant ($Myh11^{creER+};Cav1^{-/f1}$; $n = 5$ mice; 75 capillaries) mice. Statistical significance was determined by a nested unpaired, two-tailed *t*-test for **b–d, f–h**).



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Conditional SMC-specific *Cav1*-knockout mice have normal neurovascular coupling. **a**, Immunostaining on brain sections for ECs (ICAM2, green), SMCs (SMA, magenta) and CAV1 (red) from control and SMC conditional *Cav1*-mutant mice. **b**, Transmission electron microscopy images of CNS aECs and SMCs from control and SMC conditional *Cav1*-mutant mice. Arrowheads point to caveolae. L, Lumen. **c**, Mean normalized immunofluorescence of CAV1 in SMCs from control ($n = 5$ mice) and SMC-specific conditional *Cav1*-mutant mice ($n = 5$ mice). **d**, Quantification of the mean vesicular density in aECs and SMCs in control ($n = 5$ mice, 23 arterioles) and SMC conditional *Cav1*^{-/-} mice ($n = 5$ mice, 22 arterioles). **e–g**, Time course of

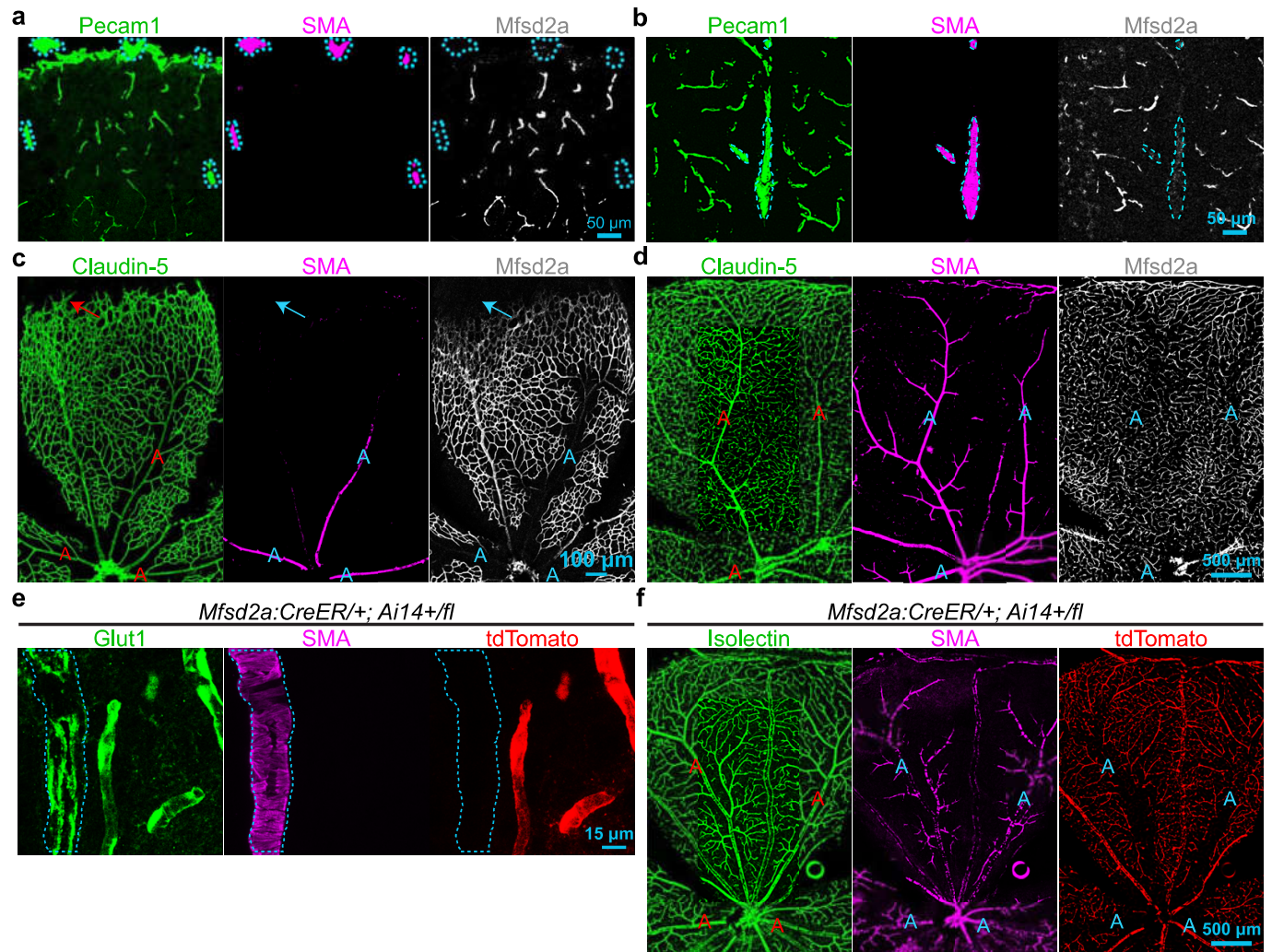
change in arteriolar dilation (**e**), maximum percentage change in arteriolar dilation (**f**) and baseline diameter (**g**) in control ($n = 7$ mice, 193 arterioles) and SMC conditional *Cav1*-mutant mice ($n = 5$ mice, 176 arterioles). **h–j**, Time course of change in red blood cell velocity (**h**), maximum percentage change in red blood cell velocity (**i**) and baseline velocity (**j**) in control ($n = 7$ mice, 75 capillaries) and SMC conditional *Cav1*-mutant mice ($n = 5$ mice, 64 capillaries). Statistical significance was determined by unpaired, two-tailed Mann–Whitney *U*-test for **c** and a nested, unpaired, two-tailed *t*-test for (**d**, **f**, **g**, **i**, **j**). Data are mean \pm s.e.m.



Extended Data Fig. 8 | See next page for caption.

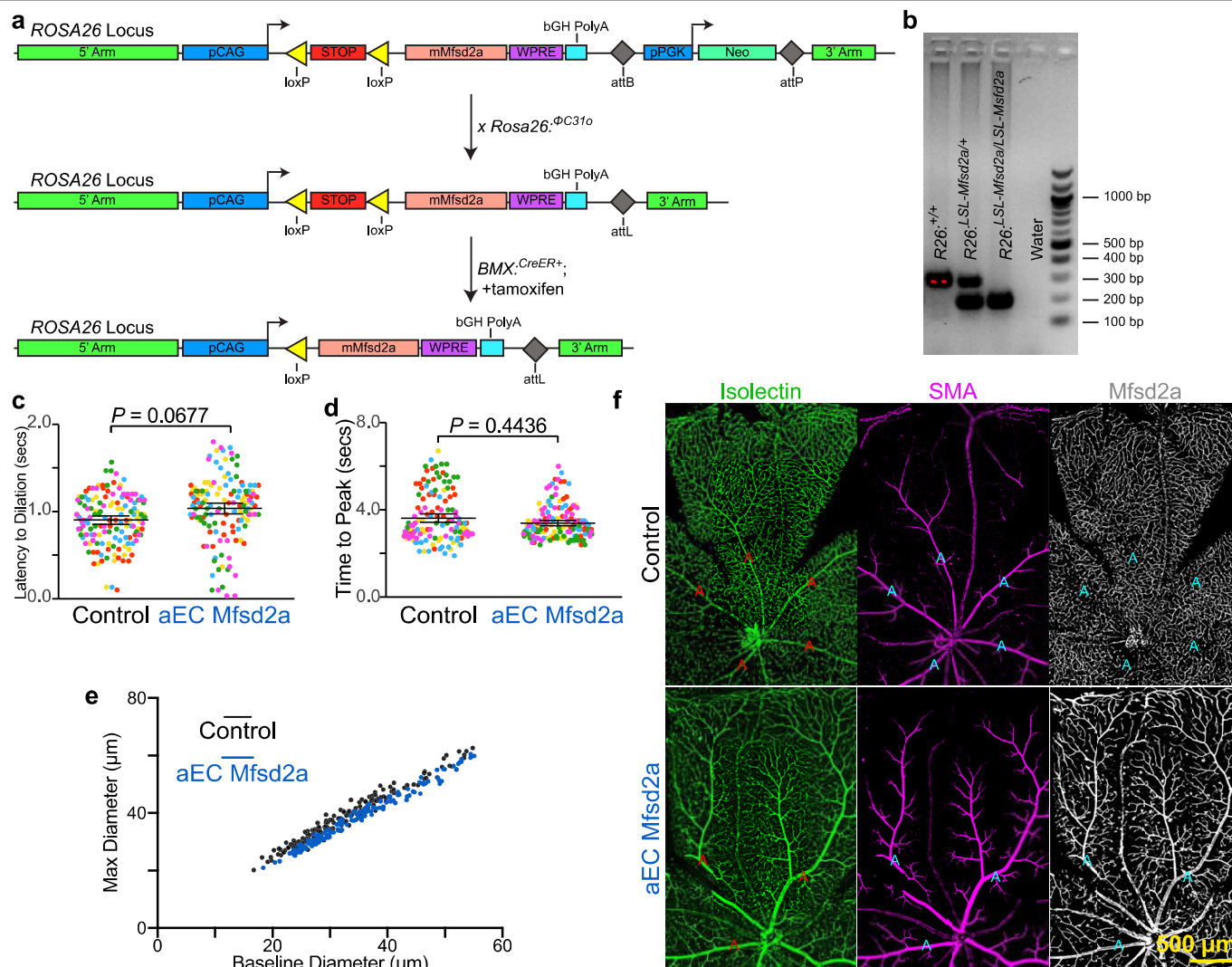
Extended Data Fig. 8 | *Cav1*-mutant mice have normal levels of eNOS protein and NO in CNS aECs and *Cav1* and *Nos3* double knockout mice have normal baseline diameter and red blood cell flow. **a**, Immunostaining on adult brain sections for ECs (PECAM1, green), arterioles (SMA, magenta) and eNOS (cyan) from *Cav1^{+/+}Nos3^{+/+}*, *Cav1^{-/-}Nos3^{+/+}* and *Cav1^{+/+}Nos3^{-/-}* mice. Independent replications were performed on three mice per genotype. **b**, Immunostaining for (PECAM1, green) and arterioles (SMA, magenta) on brain sections from *Cav1^{+/+}Nos3^{+/+}*, *Cav1^{-/-}Nos3^{+/+}* and *Cav1^{+/+}Nos3^{-/-}* mice after in vivo perfusion of NO-sensitive dye; DAF-2, yellow. Independent replicates were performed on four mice per genotype. **c**, Quantification of eNOS immunofluorescence intensity as shown in **a** in aECs from *Cav1^{+/+}Nos3^{+/+}* ($n = 3$ mice, 35 images),

Cav1^{-/-}Nos3^{+/+} ($n = 3$ mice, 35 images) and *Cav1^{+/+}Nos3^{-/-}* ($n = 3$ mice, 37 images). **d**, Quantification of DAF-2 intensity in aECs as shown in **(b)** from *Cav1^{+/+}Nos3^{+/+}* ($n = 4$ mice, 73 images), *Cav1^{-/-}Nos3^{+/+}* ($n = 4$ mice, 71 images), and *Cav1^{+/+}Nos3^{-/-}* ($n = 4$ mice, 64 images). **e, f**, Quantification of baseline diameter (**e**) and baseline velocity (**f**) in *Cav1^{+/+}Nos3^{+/+}* ($n = 5$ mice, 148 arterioles, 76 capillaries), *Cav1^{-/-}Nos3^{+/+}* ($n = 5$ mice, 128 arterioles, 68 capillaries), *Cav1^{+/+}Nos3^{-/-}* ($n = 5$ mice, 137 arterioles, 73 capillaries) and *Cav1^{-/-}Nos3^{-/-}* mice ($n = 5$ mice, 139 arterioles, 74 capillaries). Statistical significance was determined by nested, unpaired, two-tailed *t*-test for **c, d**, and nested, one-way ANOVA with a post hoc Bonferroni multiple-comparison adjustment for **e, f**. Data are mean \pm s.e.m.



Extended Data Fig. 9 | MFSD2A is not detected in CNS arterioles in brain and retina. **a, b**, Immunostaining on postnatal day (P)5 (**a**) and adult (**b**) brain sections for ECs (PECAM1, green), SMCs (SMA, magenta) and MFSD2A (white) from wild-type mice. Blue hashes outline SMA⁺ arterioles. **c, d**, Immunostaining on P5 (**c**) and adult (**d**) retina for ECs (claudin 5, green), SMCs (SMA, magenta) and MFSD2A (white) from wild-type mice. A, arterioles. MFSD2A is absent in

nascent, distal vessel (arrows) in P5 retina in **c, e, f**. Tamoxifen-treated, adult knock-in *Mfsd2a*^{CreER/+; Ai14^{+/fl} reporter mice demonstrates that tdTomato is absent in SMA⁺ arterioles but present in SMA⁺ capillaries in brain (**e**) and retina (**f**). Blue hashes and A indicate SMA⁺ arterioles. Independent replicates for **a–f** were performed on five wild-type mice.}



Extended Data Fig. 10 | Generation of a Cre-dependent MFSD2A-overexpression transgenic mouse ($R26^{LSL-Mfsd2a}$). **a**, Construct for Cre-dependent MFSD2A overexpression knocked-in to the ROSA26 locus. Mating with $ROSA26:\Phi C31$ recombinase mice removes the neomycin selection cassette. Subsequent mating with BMX^{CreER} and tamoxifen injection enables ectopic overexpression of *Mfsd2a* in aECs. **b**, PCR genotyping of Cre-dependent MFSD2A-overexpression mice. **c**, Quantification of latency to changes in arteriolar dilation in control ($BMX^{CreER-/-};R26^{LSL-Mfsd2a/+}$; $n = 5$ mice, 149 arterioles) and aEC-specific MFSD2A overexpression ($BMX^{CreER+};R26^{LSL-Mfsd2a/+}$; $n = 5$ mice; 138 arterioles) mice. **d**, Quantification of time to peak arteriolar dilation in

control ($BMX^{CreER-/-};R26^{LSL-Mfsd2a/+}$; $n = 5$ mice, 149 arterioles) and aEC-specific conditional *Cav1*-mutant ($BMX^{CreER+};R26^{LSL-Mfsd2a/+}$; $n = 5$ mice, 138 arterioles) mice. **e**, Baseline diameter to absolute maximum diameter response during whisker stimulation in control ($BMX^{CreER-/-};R26^{LSL-Mfsd2a/+}$; $n = 5$ mice, 149 arterioles) and aEC-specific conditional *Cav1*-mutant ($BMX^{CreER+};R26^{LSL-Mfsd2a/+}$; $n = 5$ mice; 138 arterioles) mice. **f**, Immunostaining on adult retinas for ECs (isolectin, green), SMCs (SMA, magenta) and MFSD2A (white) from control and aEC-specific MFSD2A-overexpression mice. Independent replications for **f** were performed on three mice per genotype. Statistical significance was determined by a nested unpaired, two-tailed *t*-test for **c**, **d**.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

The following softwares were used to collect the data in this study:

- ScanImage 5.1 for two-photon imaging (Vidrio Technologies)
- LAS X 3.0.16120.2 for Leica SP8 Confocal Imaging
- OlyVIA Ver.2.9.1 for VS120 Virtual Slide Microscope

Data analysis

- Graphpad Prism 7 and 8
- MATLAB R2018b; Script for Pial and deep Arteriolar Dilation analysis is available as listed on "Code Availability"
- Microsoft Excel 2016
- Adobe Photoshop CS6
- NIH ImageJ 2.0.0
- Geneious 10.0.09

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data for quantification mentioned in text or shown in graphs plotted in Figs. 1-4 and Extended Fig. 1-10 will be available in the online version of this paper at www.nature.com/nature. Custom MATLAB scripts and ImageJ macros used to generate and analyzed datasets in this study are available at: MATLAB script to

analyze blood flow is available from Pratik Y. Chhatbar and Prakash Kara et al. (2013).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For in vivo two photon, confocal and electron microscopy data, we performed preliminary experiments to identify the variation. We then perform a power test to identify appropriate sample sizes of images/videos per mouse. Based on previous experience with similar studies, the sample sizes were sufficient.
Data exclusions	In vivo two photon videos were excluded when the quality of the videos was too poor for data analysis. Variations in recovery from the craniotomy, in rare instances, affected the cranial window quality causing low signal to noise. Signal to noise ratio was used as a criterion for excluding two-photon data.
Replication	The experimental findings reported were verified as reproducible through the collection of multiple technical replicates and mice. When possible, the data was collected and analyzed blindly to genotype.
Randomization	Mice were randomized based on their genotypes and allocated randomly in the their respective genotype group. Field of views were randomly sampled.
Blinding	Acquisition, collection and analysis of the experiments were performed all blinded to the genotypes and performed by independent people. Only after the data was completely analyzed were the genotypes unblinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Antibodies

Antibodies used	<p> α-Mfsd2a (1:200, Cell Signaling Technologies; RRID: AB_2617168) α-Mfsd2a (generous gift from Dr. David Silver as used previously) α-SMA (1:1000, Sigma-Aldrich Cat# C6198, RRID:AB_476856; Clone name:1A4; lot no.: 058M4761V) α-ICAM2 (1:200, BD Biosciences Cat# 553326, RRID:AB_394784; Clone 3C4(mIC2/4); lot no.: 8248688) α-PECAM (R and D Systems Cat# AF3628, RRID:AB_2161028) α-Claudin-5 (Thermo Fisher Scientific Cat# 34-1600, RRID:AB_2533157, lot no.: TJ276183) α-eNOS (Abcam Cat# ab5589, RRID:AB_304967) α-Cav1 (Cell Signaling Technology Cat# 3267, RRID:AB_2275453, Clone D46G3) α-Glut1 (Millipore Cat# 07-1401, RRID:AB_1587074; lot 3021918) α-CNN1 (Abcam Cat# ab46794, RRID:AB_2291941, Clone EP798Y) α-Desmin (Thermo Fisher Scientific Cat# PA1-37556, RRID:AB_2292917; lot no.: Q12081282) α-Tagln (Abcam Cat# ab14106, RRID:AB_443021) α-Myh11 (Abcam Cat# ab53219, RRID:AB_2147146; lot no.: GR3261273-2) </p>
Validation	<p> α-Mfsd2a (Cell Signaling) has been validated using immunohistochemistry on Mfsd2a brain KO tissue as previously demonstrated in Ben-Zvi et al. Nature 2014 α-Mfsd2a (Gift from David Silver) has been validated using immunohistochemistry on Mfsd2a brain KO tissue as previously </p>

demonstrated in Nyguen et al. Nature 2014.

α -SMA is valid using immunohistochemistry because staining was specifically localized to smooth muscle cells on arteries.

α -ICAM2 is valid using immunohistochemistry because the staining is consistent with the known expression of endothelial cells and manufacturer routinely test by flow cytometry.

α -PECAM is valid using immunohistochemistry because the staining is consistent with the known expression of endothelial cells and extensively used in many publications.

α -Claudin-5 is valid using immunohistochemistry because the staining is consistent with the known expression of CNS endothelial cells and extensively used in many publications

α -eNOS is valid using immunohistochemistry because antibody has been validated on eNOS knockout tissue on brain and expression is consistent with endothelial cells

α -Cav1 is valid using immunohistochemistry because antibody has been validated on Cav1 knockout tissue as previously validated on Andreone et al. 2017. Neuron

α -Glut1 is valid using immunohistochemistry because the staining is consistent with the known expression of CNS endothelial cells

α -CNN1 is valid using immunohistochemistry because staining was specifically localized to smooth muscle cells on arteries and used previously by (Vanlandewijck et al., 2018).

α -Desmin is valid using immunohistochemistry because the staining is consistent with the known expression of mural cells.

α -Tagln is valid because staining was specifically localized to smooth muscle cells on arteries and used previously by (Vanlandewijck et al., 2018).

α -Myh11 is valid using immunohistochemistry because staining was specifically localized to smooth muscle cells on arteries and used previously by (Vanlandewijck et al., 2018).

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

All animal experiments were approved by the Harvard University Institutional Animal Care and Use Committee (IACUC). The following mice strains were used: wildtype (C57BL/6J, Jackson Laboratory # 000664), Mfsd2a:CreER(Q. Chen et al., 2016), BMX:CreER(Ehling et al., 2013), Myh11:CreER(Wirth et al., 2008), NG2:DsRED(Zhu et al., 2008) JAX# 008241, Ai14 (Madisen et al., 2010) JAX# 007914, Ai39(Madisen et al., 2012) JAX# 014539, Ai75(Quina et al., 2017) JAX# 014539, ROSA26:LSL-Mfsd2a (this paper), Thy1:GCaMP6s(T.-W. Chen et al., 2013) JAX# 024275, Cav1^{-/-} (Razani et al., 2001) JAX# 007083, Cav1 floxed(Asterholm et al., 2012), Nos3^{-/-} (Shesely et al., 1996) JAX# 002684, ROSA26: PhiC31(Raymond and Soriano, 2007) JAX# 007743. Mice of both sexes were used and ages ranged from 8 to 20 weeks old.

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve field-collected samples.

Ethics oversight

All animal experiments were approved by the Harvard University Institutional Animal Care and Use Committee (IACUC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Lipid availability determines fate of skeletal progenitor cells via SOX9

<https://doi.org/10.1038/s41586-020-2050-1>

Received: 1 July 2016

Accepted: 8 January 2020

Published online: 26 February 2020

 Check for updates

Nick van Gestel^{1,2,3,4,5}, Steve Stegen^{1,2}, Guy Eelen^{6,7}, Sandra Schoors^{6,7}, Aurélie Carlier^{2,8,9,10}, Veerle W. Daniëls^{11,12}, Ninib Baryawno^{3,4,5,13}, Dariusz Przybylski¹⁴, Maarten Depypere^{15,16}, Pieter-Jan Stiers^{1,2}, Dennis Lambrechts^{2,17,18}, Riet Van Loooveren¹, Sophie Torrekens¹, Azeem Sharda^{3,4,5}, Patrizia Agostinis¹⁹, Diether Lambrechts^{20,21}, Frederik Maes^{15,16}, Johan V. Swinnen¹¹, Liesbet Geris^{2,8,9}, Hans Van Oosterwyck^{2,8}, Bernard Thienpont^{20,21}, Peter Carmeliet^{6,7}, David T. Scadden^{3,4,5} & Geert Carmeliet^{1,2}✉

The avascular nature of cartilage makes it a unique tissue^{1–4}, but whether and how the absence of nutrient supply regulates chondrogenesis remain unknown. Here we show that obstruction of vascular invasion during bone healing favours chondrogenic over osteogenic differentiation of skeletal progenitor cells. Unexpectedly, this process is driven by a decreased availability of extracellular lipids. When lipids are scarce, skeletal progenitors activate forkhead box O (FOXO) transcription factors, which bind to the *Sox9* promoter and increase its expression. Besides initiating chondrogenesis, SOX9 acts as a regulator of cellular metabolism by suppressing oxidation of fatty acids, and thus adapts the cells to an avascular life. Our results define lipid scarcity as an important determinant of chondrogenic commitment, reveal a role for FOXO transcription factors during lipid starvation, and identify SOX9 as a critical metabolic mediator. These data highlight the importance of the nutritional microenvironment in the specification of skeletal cell fate.

Bone repair reiterates the developmental endochondral ossification process and is initiated by periosteal skeletal progenitor cells that first form an avascular cartilage template which is later replaced by bone^{1,2}. Among the factors involved in chondrogenesis, the transcription factor SOX9 has been the most extensively studied, but how it is induced in skeletal progenitor cells is poorly understood. Since cartilage is avascular, the absence of blood vessels itself has been suggested to initiate chondrogenesis^{3–6}, but a causal link has not been confirmed and remains controversial⁷. In this study, we provide evidence that local blood vessel availability determines skeletal progenitor cell fate during bone healing through a multifaceted mechanism involving lipid metabolism, FOXO signalling and SOX9.

Vascularity controls skeletal cell fate

To investigate whether the absence of vasculature determines skeletal progenitor fate we transplanted viable (autologous) bone grafts into femoral defects in mice, inducing a periosteal-driven healing

response⁸. Periosteal progenitor cells near the host-graft border formed cartilage, whereas cells in the centre differentiated directly into bone-forming osteoblasts (Extended Data Fig. 1a). Periosteal cells did not contribute to blood vessels in the callus (Fig. 1a), but actively promoted vascular ingrowth as their removal reduced bone formation and callus vascularization (Extended Data Fig. 1b–d). At post-fracture day (PFD) 7 the central periosteal callus vasculature was highly connected with that of the surrounding muscle (Fig. 1b), suggesting that periosteal cells attract blood vessels from this site. To investigate the importance of this vascular ingrowth for bone repair, we inserted polycarbonate filters with different pore sizes between graft and muscle (Fig. 1c). Inserting a 30- μ m-pore filter still allowed capillaries to transverse the pores at PFD7, whereas a 0.2- μ m-pore size prevented vascular ingrowth into the periosteal layer, as shown by the numerous capillaries adjacent to the filter on the muscle side and reduced callus vascularization (Fig. 1d). Concomitantly, periosteal cellularity decreased because of reduced proliferation and moderately increased cell death (Extended Data Fig. 2a, b), but more importantly,

¹Laboratory of Clinical and Experimental Endocrinology, Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, Leuven, Belgium. ²Prometheus, Division of Skeletal Tissue Engineering, KU Leuven, Leuven, Belgium. ³Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. ⁴Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA. ⁵Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA. ⁶Laboratory of Angiogenesis and Vascular Metabolism, Department of Oncology, KU Leuven, Leuven, Belgium. ⁷Laboratory of Angiogenesis and Vascular Metabolism, Center for Cancer Biology, VIB, Leuven, Belgium. ⁸Biomechanics Section, Department of Mechanical Engineering, KU Leuven, Leuven, Belgium. ⁹Biomechanics Research Unit, GIGA In Silico Medicine, University of Liège, Liège, Belgium. ¹⁰MERLN Institute of Technology-Inspired Regenerative Medicine, Maastricht University, Maastricht, The Netherlands. ¹¹Laboratory of Lipid Metabolism and Cancer, Department of Oncology, KU Leuven, Leuven, Belgium. ¹²Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. ¹³Childhood Cancer Research Unit, Department of Children's and Women's Health, Karolinska Institutet, Stockholm, Sweden. ¹⁴Howard Hughes Medical Institute and Department of Biology, Brandeis University, Waltham, MA, USA. ¹⁵Medical Imaging Research Center, KU Leuven, Leuven, Belgium. ¹⁶Department of Electrical Engineering, ESAT/PSI, Medical Image Computing, KU Leuven, Leuven, Belgium. ¹⁷Centre for Surface Chemistry and Catalysis, Department of Microbial and Molecular Systems, KU Leuven, Leuven, Belgium. ¹⁸Tissue Engineering Laboratory, Skeletal Biology and Engineering Research Center, Department of Development and Regeneration, KU Leuven, Leuven, Belgium. ¹⁹Laboratory of Cell Death Research and Therapy, Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium. ²⁰Laboratory of Translational Genetics and Laboratory for Functional Epigenetics, Department of Human Genetics, KU Leuven, Leuven, Belgium. ²¹Laboratory of Translational Genetics, Center for Cancer Biology, VIB, Leuven, Belgium. ✉e-mail: geert.carmeliet@kuleuven.be

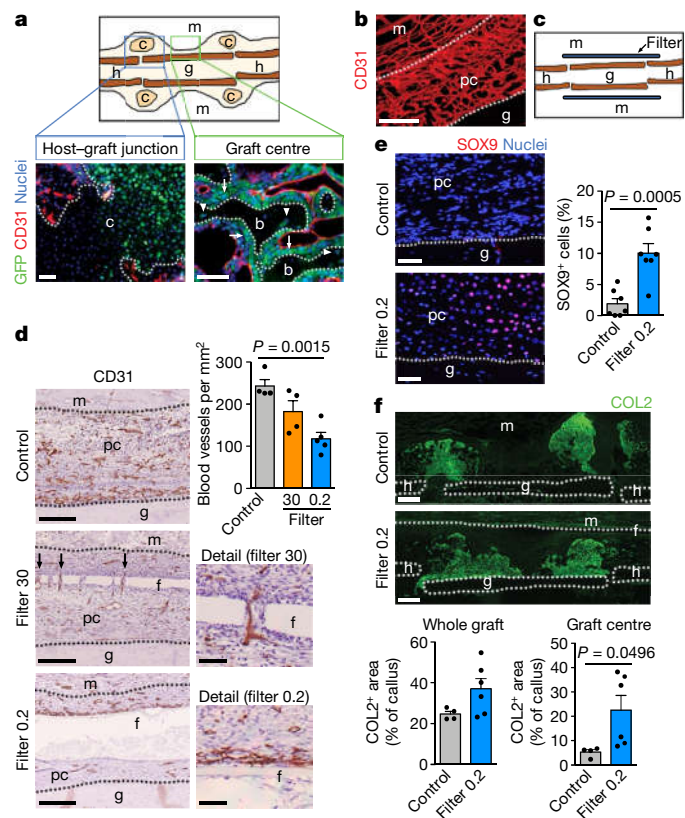


Fig. 1 | Preventing vascular ingrowth during bone healing induces chondrogenesis. **a**, Immunofluorescence analysis of bone-graft periosteal cell tracing showing contribution to cartilage and bone (arrows, GFP⁺ osteoblasts; arrowheads, GFP⁺ osteocytes) in the graft callus at PFD14, while CD31⁺ blood vessels (red) are mainly host-derived (representative images of 4 mice). Scale bars, 50 μ m. **b**, Immunofluorescence analysis of a bone-autograft section revealing the interconnected periosteal callus and skeletal muscle vasculature at PFD7 (representative image of 3 mice). Scale bar, 200 μ m. **c**, Schematic representation of the autograft model with filter. **d**, Immunohistochemical analysis and quantification of callus vascularization at PFD7 when a filter with 30 μ m (filter 30; arrows indicate blood vessels passing through filter pores) or 0.2 μ m (filter 0.2) pore size was placed in between muscle and graft ($n = 4$ mice for control and filter 30, $n = 5$ mice for filter 0.2). Scale bars, 50 μ m in detail images, otherwise 200 μ m. **e**, Visualization and quantification of early chondrogenic cells in the callus of grafts with or without a filter (0.2 μ m) at PFD7 by immunofluorescence for SOX9 ($n = 7$ mice). Scale bars, 50 μ m. **f**, Visualization and quantification of cartilage in the callus of autografts with and without filter (0.2 μ m) at PFD14 by immunofluorescence for collagen type 2 (COL2) ($n = 4$ mice for control, $n = 6$ mice for filter 0.2). Scale bars, 500 μ m. b, bone; c, cartilage; f, filter; g, graft; h, host; m, muscle; pc, periosteal callus. Data are mean \pm s.e.m.; one-way ANOVA with Bonferroni post hoc test (**d**), two-tailed Student's *t*-test (**e**, **f**).

the number of SOX9⁺ early chondrogenic cells was higher at the central graft region (Fig. 1e). This chondrogenic switch resulted in less bone (Extended Data Fig. 2c) but more type 2 collagen (COL2)⁺ cartilage matrix in the central region at PFD14 (Fig. 1f), where graft cells differentiated to chondrocytes instead of osteoblasts (Extended Data Fig. 2d). At PFD28, successful healing was observed in both conditions, although the presence of small cartilage islands in the callus with filter ($75.0 \pm 14.4\%$ of sections) suggests delayed healing (Extended Data Fig. 2e, f). Thus, skeletal progenitor cells undergo chondrogenic rather than osteogenic differentiation when blood supply is limited, securing successful bone healing.

During bone healing, the vasculature supplies nutrients (oxygen, glucose, amino acids and lipids), growth factors and perivascular

progenitor cells⁹. To distinguish between these components, we applied a computational model of bone healing^{10,11} to our bone-graft setup, in which cell fate and tissue formation are controlled by nutrient availability, osteochondrogenic growth factors, matrix density and local cell number (Extended Data Fig. 3a, b). The model correctly described the spatiotemporal progression of normal bone-graft healing (that is, blood vessels can come from the muscle; compare Extended Data Fig. 3b with Extended Data Fig. 1a). When the presence of a filter was mimicked by limiting diffusion of nutrients from the muscle side (20–40% of the nutrients normally supplied by the vasculature), the model recapitulated the chondrogenic switch in the central graft region (Extended Data Fig. 3c, d). An additional supply of growth factors and/or progenitor cells from the muscle side did not significantly affect this bone repair profile (Extended Data Fig. 3e). The in silico model thus supports the hypothesis that nutrients supplied by the vasculature regulate skeletal progenitor cell differentiation.

Lipid scarcity induces chondrogenesis

To test this hypothesis, we investigated the nutritional control of cell fate using two models of skeletal progenitors: the C3H10T1/2 cell line, a homogeneous population retaining multipotency properties¹², and primary murine periosteal cells, which are more heterogeneous but contain true skeletal stem and progenitor cells^{13–15}. We confirmed key findings in immunophenotypically-defined skeletal stem cells isolated from total long bones of newborn mice¹⁶, which are homogeneous but limited in number.

Combined nutrient deprivation (CND; reduced levels of serum, oxygen, glucose and glutamine) increased SOX9 protein and mRNA levels in C3H10T1/2 or periosteal cells, without changes in expression of osteogenic, adipogenic or myogenic transcription factors (Fig. 2a, Extended Data Fig. 4a–c). Depriving C3H10T1/2 cells of individual nutrients revealed that low oxygen levels increased SOX9, as reported^{17,18}, whereas lowering glucose or glutamine levels had little effect (Fig. 2b). Unexpectedly, serum deprivation led to massive and rapid accumulation of SOX9 mRNA and protein, resulting from increased transcription and translation (Fig. 2b, Extended Data Fig. 4d–g). Expression of osteogenic, adipogenic and myogenic transcription factors did not change (Extended Data Fig. 4h). Serum deprivation also increased SOX9 in periosteal cells (Extended Data Fig. 4i) and enhanced their chondrogenic differentiation in micromass cultures (Fig. 2c), but prevented osteogenic differentiation (Extended Data Fig. 4j). A possible explanation for this chondrogenic switch is avoiding cell death. Indeed, knockdown of SOX9 in C3H10T1/2 cells, periosteal cells and growth plate-derived chondrocytes reduced cell viability in CND, and to a minor extent also in serum deprivation (Extended Data Fig. 4k, l). Thus, skeletal progenitor cells rapidly adapt to specific nutritional stress by increasing SOX9 levels and undergoing chondrogenic commitment.

Serum represents the main source of lipids, and we questioned whether serum deprivation-induced chondrogenic commitment of skeletal progenitors could be attributed to lipid deprivation. Resupplying C3H10T1/2 cells with oleate (Fig. 2d), palmitate, very low density lipoproteins or polyunsaturated fatty acids (PUFA) (Extended Data Fig. 5a–c) prevented the increase in SOX9 during serum deprivation. In addition, lipid-reduced serum (LRS) mimicked the effects of serum deprivation. LRS increased SOX9 levels in C3H10T1/2 cells (Fig. 2e), promoted chondrogenic differentiation of periosteal cells in micromass or pellet cultures, an effect partially reversed by exogenous fatty acids (Fig. 2f, Extended Data Fig. 5d), and inhibited their osteogenic differentiation (Extended Data Fig. 5e). Importantly, serum deprivation or LRS also increased SOX9 levels in skeletal stem cells (Extended Data Fig. 5f). In all studied cell types, lipid deprivation increased the number of SOX9^{high} cells, and cell cycle and apoptosis analysis showed this was

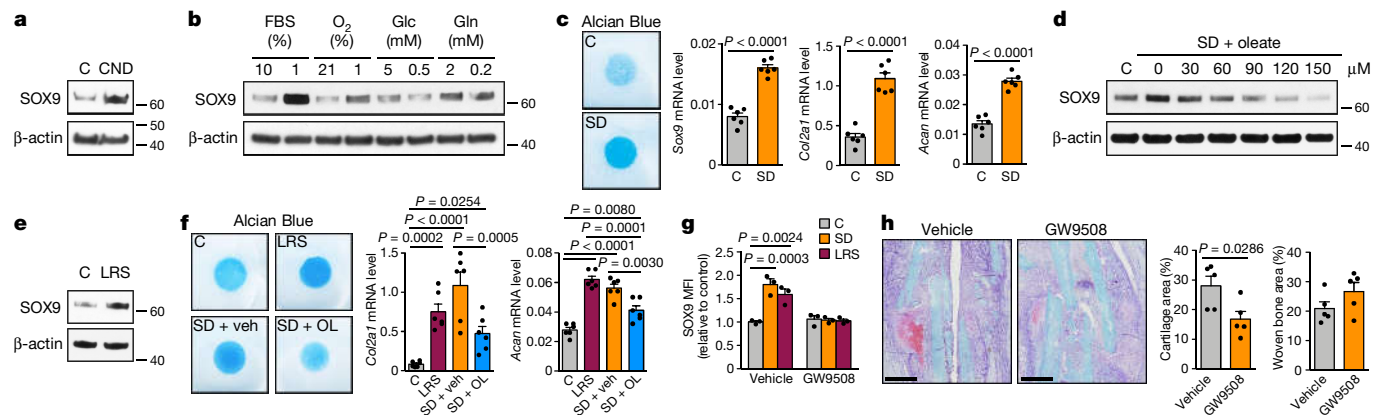


Fig. 2 | Lipid scarcity induces SOX9 in skeletal progenitors. **a, b**, Immunoblot detection of total SOX9 in C3H10T1/2 cells exposed for 24 h to control (C) or CND medium (a) or to different nutritional stresses (FBS, fetal bovine serum; Glc, glucose; Gln, glutamine) (b), with β -actin as loading control ($n = 2$ independent experiments). **c**, Chondrogenic differentiation of periosteal cells in control or serum deprivation (SD) medium, assessed by visualization of chondrogenic matrix deposition (alcian blue staining) and quantification of *Sox9*, *Col2a1* and *Acan* mRNA levels (relative to *Actb*, $n = 6$ biologically independent samples). **d, e**, Immunoblot detection of total SOX9 in C3H10T1/2 cells exposed for 6 h to control medium, SD medium, SD medium supplemented with increasing concentrations of oleate (d) or LRS medium (e), with β -actin as loading control ($n = 2$ independent experiments). **f**, Chondrogenic differentiation of periosteal cells in control medium, LRS

medium, SD medium or SD medium supplemented with 60 μ M oleate (OL), assessed by alcian blue staining and quantification of *Col2a1* and *Acan* mRNA levels (relative to *Actb*, $n = 6$ biologically independent samples). Veh, vehicle. **g**, Flow cytometric quantification of total SOX9 levels in periosteal cells exposed for 24 h to control medium, SD medium or LRS medium supplemented with 100 μ M GW9508 (FFAR1/4 agonist) or vehicle (DMSO) ($n = 3$ biologically independent samples). **h**, Histological visualization (safranin O staining) and quantification of cartilage and woven bone in the callus at PF07 of mice treated daily with GW9508 (10 nmol) or vehicle (0.2% DMSO in saline) at the fracture site ($n = 5$ mice). Scale bars, 500 μ m. Data are mean \pm s.e.m.; two-tailed Student's *t*-test (c, h), one-way ANOVA (f) or two-way ANOVA (g) with Bonferroni post hoc test. For gel source data, see Supplementary Fig. 1.

not due to selection of a pre-existing SOX9^{high} population (Extended Data Fig. 5f–h).

We next tested whether lipid availability also controls skeletal progenitor differentiation in more physiologically relevant settings. Since it is not feasible to locally deprive cells specifically of exogenous lipids in vivo, we first used embryonic metatarsal cultures, an organ-like ex vivo model of bone development. Serum deprivation increased the number of SOX9⁺ chondrocytes and prevented osteogenesis, evidenced by absence of Col1a1-expressing cells and mineralization, which was reversed by fatty acid supplementation (Extended Data Fig. 5i, j). Second, local injection of fatty acids during fracture repair reduced the amount of cartilage in the callus, with no change in newly formed bone (Extended Data Fig. 5k). Third, GW9508, an agonist of free fatty acid receptor 1 (FFAR1) and FFAR4, prevented the increase in SOX9 induced by serum deprivation or LRS in the three cell models (Fig. 2g; Extended Data Fig. 5l). Accordingly, locally injecting GW9508 during fracture repair decreased cartilage in the callus without affecting woven bone areas (Fig. 2h). Together, low local lipid levels promote chondrogenesis of skeletal progenitor cells in vivo.

Our findings suggest that the chondrogenic switch during bone-graft healing in the presence of a filter (Fig. 1) is primarily due to the absence of exogenous lipids, which is linked to poor vascularization. We found that diffusion of lipids in a collagen gel containing periosteal cells is much lower than that of glucose (Extended Data Fig. 5m), indicating that lipids are a limiting nutrient when vascularization is inadequate. Furthermore, we showed that the absence of specific cell types, potentially blocked by the filter, does not impact chondrogenesis. Indeed, serum deprivation-supported chondrogenic differentiation of periosteal cells in micromass cultures was not prevented by muscle-derived endothelial cells, macrophages or pericytes, in contrast to fatty acid supplementation. (Extended Data Fig. 5n, o). Together with our in vivo (Fig. 1) and in silico (Extended Data Fig. 3) results, this shows that lipid deprivation caused by reduced vascularization is probably an important determinant of periosteal chondrogenesis during bone healing.

Chondrocytes have low fatty acid oxidation

Why would chondrogenic commitment be beneficial when lipids are scarce? We hypothesized that chondrocyte metabolism does not rely on exogenous lipids. To test this, we compared the metabolic profile of chondrocytes to that of skeletal progenitors and mature osteoblasts (Fig. 3a, Extended Data Fig. 6a). Chondrocytes were highly glycolytic, as reported^{19,20}. Osteoblasts showed the highest oxygen consumption rate (OCR), which was not owing to high glucose oxidation, but to a higher rate of fatty acid oxidation (FAO). Chondrocytes exhibited low FAO and skeletal progenitors had an intermediate profile. To confirm these findings in vivo, we examined metabolic-gene signatures in a mouse long-bone single-cell RNA-sequencing (RNA-seq) dataset that we generated recently²¹. This atlas encompasses 17 non-haematopoietic cell types including skeletal progenitors, chondrocytes and osteoblasts (Extended Data Fig. 6b). The different chondrocyte populations (clusters 2, 10, 13, 17; *Sox9*⁺*Acan*⁺) showed low expression of FAO genes and high expression of glycolytic genes compared with osteoblasts (clusters 7 and 8; *Col1a1*⁺*Ocn*⁺; *Ocn* is also known as *Bglap*) and, to a minor extent, skeletal progenitors (clusters 1 and 4; *Grem1*⁺) (Extended Data Fig. 6b, c). Gene expression analysis confirmed higher expression of the glycolytic genes *Slc2a1* (encoding GLUT1), *Pfkfb3* and *Ldha*, but lower expression of the FAO-related genes *Cpt1a*, *Acadm* and *Acadl* in growth plate cartilage versus cortical bone samples (Extended Data Fig. 6d). Immunohistochemistry showed low CPT1a levels and high GLUT1 levels in chondrocytes of the growth plate and fracture callus, whereas trabecular bone osteoblasts displayed high levels of both CPT1a and GLUT1 (Fig. 3b). Intravenous injection of fluorescent fatty acid and glucose analogues revealed labelled fatty acids in osteoblasts but not in chondrocytes in the growth plate or fracture callus, whereas labelled glucose was taken up by both cell types (Extended Data Fig. 6e, f), confirming that low FAO in chondrocytes correlates with lipid scarcity. Transplantation experiments showed that loss of CPT1a abrogates osteogenic differentiation of skeletal stem cells during fracture healing but preserves their ability to become chondrocytes (Fig. 3c, Extended Data Fig. 6g). In addition, etomoxir, a CPT1 inhibitor, decreased viability

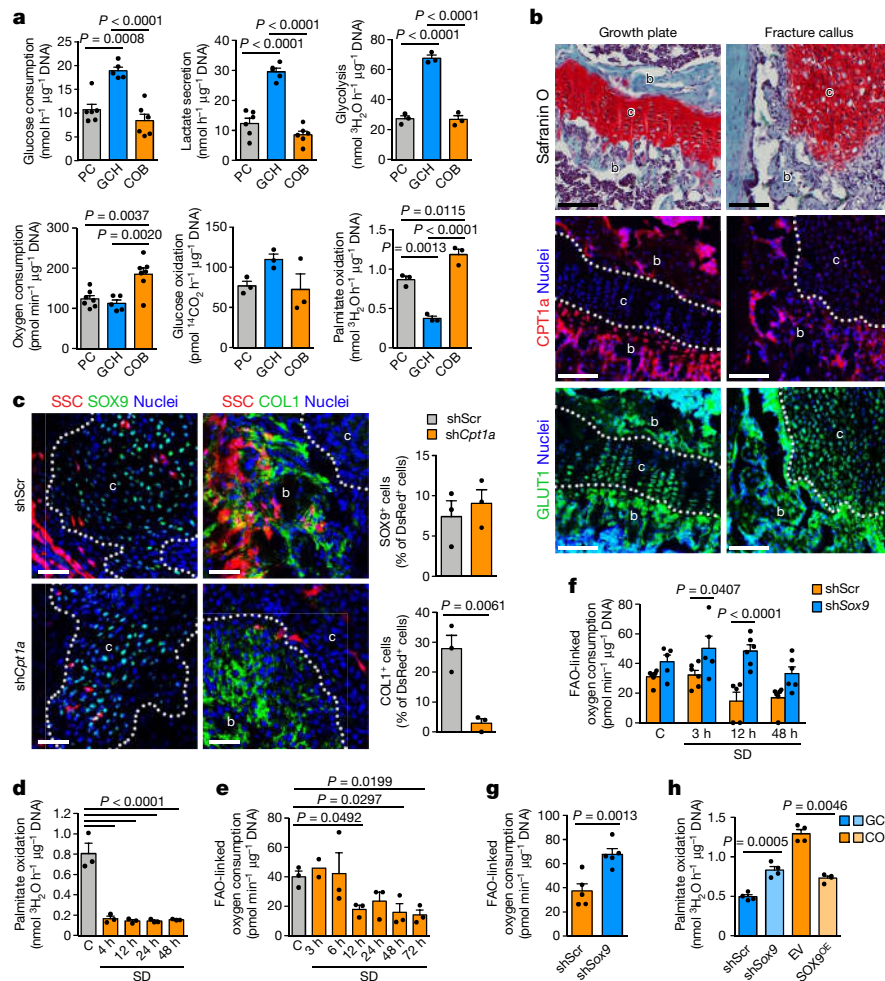


Fig. 3 | SOX9 suppresses FAO in chondrocytes. **a**, Quantification of glucose consumption and lactate secretion (PC, COB; $n=6$; GCH; $n=5$ biologically independent samples), glycolytic rate ($n=3$ biologically independent samples), oxygen consumption (PC, COB; $n=7$; GCH; $n=5$ biologically independent samples), glucose oxidation ($n=3$ biologically independent samples) and palmitate oxidation ($n=3$ biologically independent samples) in periosteal cells (PC), growth plate-derived chondrocytes (GCH) and calvarial osteoblasts (COB). **b**, Analysis of adjacent histological sections of a growth plate and fracture callus (PFD7) by safranin O staining (cartilage) or immunofluorescence for CPT1a or GLUT1 (representative images of 3 mice). Scale bars, 100 μm . Dotted white lines delineate cartilage areas. **c**, Histological visualization and quantification of early chondrogenic (SOX9⁺) and osteogenic (COL1⁺) cells in the callus of fractures (PFD7) transplanted with CAG-DsRed⁺ skeletal stem cells (SSC) transduced with shRNA against *Cpt1a* (sh*Cpt1a*) or scrambled shRNA control (shScr) ($n=3$ mice). Scale bars, 50 μm .

d, Measurement of oxidation of extracellularly added palmitate by periosteal cells in control medium or at different times in SD medium ($n=3$ biologically independent samples). **e**, Quantification of FAO-linked OCR in periosteal cells in control medium or at different times in SD medium (3h: $n=2$, other time points: $n=3$ biologically independent samples). **f**, Quantification of FAO-linked OCR in periosteal cells transduced with sh*Sox9* or shScr in control medium or at different times in SD medium (shScr 12 h, sh*Sox9* control, sh*Sox9* 3 h: $n=5$; all others: $n=6$ biologically independent samples). **g**, Quantification of FAO-linked OCR in GCH transduced with sh*Sox9* or shScr ($n=5$ biologically independent samples). **h**, Quantification of palmitate oxidation in GCH transduced with sh*Sox9* or shScr, and in COB transduced with a lentiviral vector encoding SOX9 (SOX9^{OE}, SOX9 overexpression) or an empty vector (EV) ($n=4$ biologically independent samples). Data are mean \pm s.e.m.; one-way ANOVA (**a**, **d**, **e**) or two-way ANOVA (**f**) with Bonferroni post hoc test, two-tailed Student's *t*-test (**c**, **g**, **h**).

and numbers of cultured calvarial osteoblasts but not growth plate-derived chondrocytes (Extended Data Fig. 6h). Thus, chondrocytes exhibit a low rate of FAO consistent with local lipid scarcity, and do not depend on this pathway to fulfil their metabolic demands.

SOX9 suppresses FAO

We next determined how lipid deprivation affects the rate of FAO in skeletal progenitor cells. As expected, oxidation of extracellular palmitate immediately dropped after exposing periosteal cells to serum deprivation or LRS (Fig. 3d, Extended Data Fig. 7a). Surprisingly, cells temporarily maintained total FAO, which was quantified indirectly by measuring etomoxir-sensitive OCR²², for 6 h after serum deprivation (Fig. 3e, Extended Data Fig. 7b), suggesting that they initially

compensate for the scarcity of exogenous lipids, possibly through mobilization of intracellular lipid stores. Indeed, fluorescent fatty acids translocated from lipid droplets into mitochondria, where FAO takes place, when periosteal cells were exposed to serum deprivation (Extended Data Fig. 7c). Starvation-induced lipid-droplet generation and breakdown are both linked to autophagy^{23,24}, and we confirmed that C3H10T1/2 cells and periosteal cells activate autophagy early after serum deprivation (Extended Data Fig. 7d–f). Accordingly, lipid-droplet number initially increased during serum deprivation in C3H10T1/2 cells before decreasing at 6 h, and knockdown of the essential autophagosome protein ATG5²⁵ prevented both the initial increase and the late breakdown of lipid droplets after serum deprivation (Extended Data Fig. 7g). Furthermore, the lysosomotropic drug chloroquine immediately reduced the FAO-linked OCR upon exposure

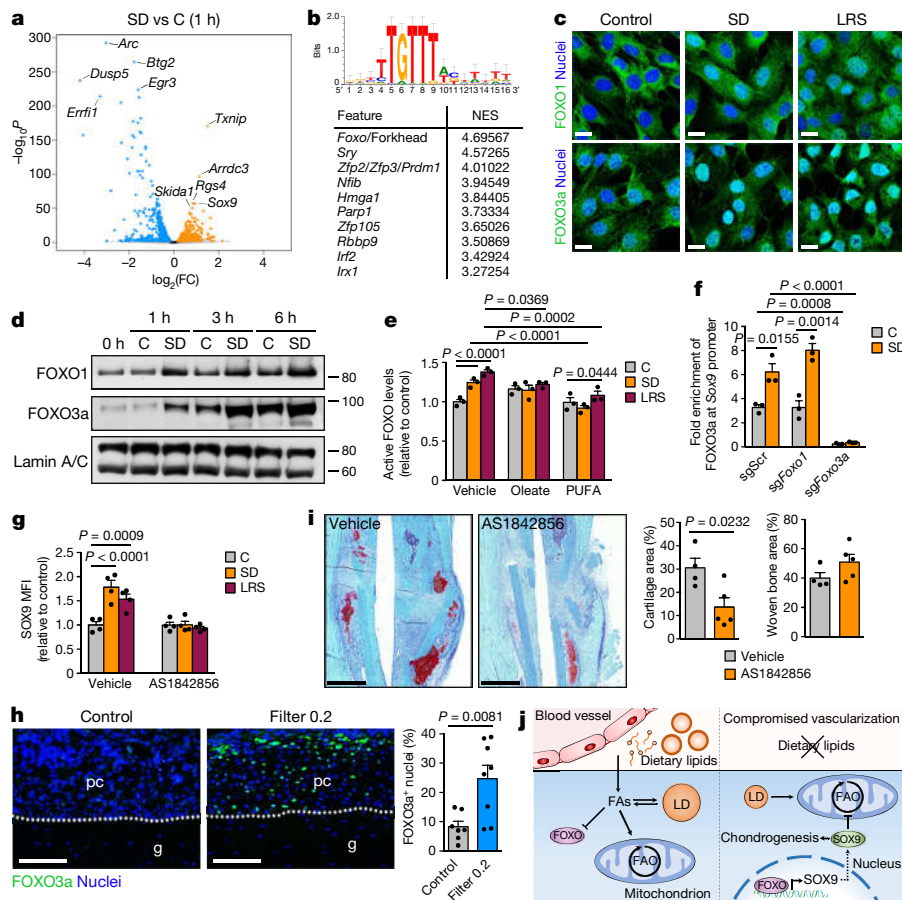


Fig. 4 | Lipids regulate SOX9 through FOXO signalling. **a**, **b**, Volcano plot showing significantly enriched and depleted mRNAs (**a**) and top-10 most significantly enriched transcription factor motifs with normalized enrichment scores (NES) as determined by i-cisTarget analysis (**b**) in C3H10T1/2 cells exposed for 1 h to SD medium versus control medium ($n = 3$ replicates). Motif shown on top is the FOXO/forkhead motif. **c**, Confocal microscopy of C3H10T1/2 cells stained for FOXO1 (top) or FOXO3a (bottom) shows increased nuclear localization after exposure of cells for 3 h to SD medium or LRS medium (representative images of 2 independent experiments). Scale bars, 20 μ m. **d**, Immunoblot detection of nuclear FOXO1 and FOXO3a in C3H10T1/2 cells exposed for 1, 3 or 6 h to control medium or SD medium, with lamin A/C as loading control ($n = 2$ independent experiments). **e**, Nuclear FOXO activity in C3H10T1/2 cells exposed for 3 h to control medium, SD medium or LRS medium supplemented with vehicle (EtOH), oleate (60 μ M) or PUFA ($n = 3$ independent experiments). **f**, Occupancy of FOXO3a at the *Sox9* promoter of Cas9-expressing C3H10T1/2 cells transduced with inducible short guide RNA (sgRNA) against *FoxO1* (sg*FoxO1*), sg*FoxO3a* or a scrambled sgRNA (sgScr), exposed for 3 h to control medium or SD medium in the presence of doxycycline (250 ng ml⁻¹), as determined by chromatin immunoprecipitation with quantitative PCR (ChIP-qPCR) ($n = 3$ independent experiments). **g**, Flow cytometric quantification of total SOX9 levels in periosteal cells exposed for

24 h to control medium, SD medium or LRS medium supplemented with 1 μ M AS1842856 (FOXO inhibitor) or vehicle (DMSO) ($n = 4$ biologically independent samples). **h**, Histological visualization and quantification of FOXO3a-expressing cells in the central periosteal callus of grafts with or without a filter (0.2 μ m pore size) at PFD7 (control: $n = 7$, filter 0.2: $n = 8$ mice). Scale bars, 50 μ m. **i**, Histological visualization (safranin O staining) and quantification of cartilage and woven bone in the callus at PFD7 of mice treated daily with AS1842856 (500 pmol) or vehicle (0.1% DMSO in saline) at the fracture site (vehicle: $n = 4$, AS1842856: $n = 5$ mice). Scale bars, 500 μ m. **j**, Schematic overview of main findings. During bone fracture healing, skeletal progenitor cells in adequately vascularized environments differentiate into osteoblasts, which require high levels of exogenous fatty acids (FA) to fuel their FAO-dependent metabolism. Cells in regions with a poor vascular supply will temporarily support FAO by breaking down intracellular lipid droplets (LD), while the lack of dietary lipids simultaneously increases FOXO activity. FOXOs increase levels of SOX9, which activates the chondrogenic program and blocks FAO. These adaptations promote cell survival and secure bone healing even in nutrient-poor environments. Data are mean \pm s.e.m.; two-way ANOVA with Bonferroni post hoc test (**e–g**), two-tailed Student's *t*-test (**h, i**). For gel source data, see Supplementary Fig. 1.

of periosteal cells to serum deprivation (Extended Data Fig. 7h) and decreased survival of C3H10T1/2 cells and periosteal cells during serum deprivation (Extended Data Fig. 7i). Together, these data show that skeletal progenitors depend on lysosome-mediated mobilization of intracellular lipid stores to temporarily support FAO and secure survival when extracellular lipids become limited.

The increase in SOX9 levels (Extended Data Fig. 4d, e) and the decrease in total FAO (Fig. 3e) occur concomitantly after lipid deprivation, suggesting that they are connected. Deletion of SOX9 in periosteal cells prevented the suppression of FAO by serum deprivation (Fig. 3f), whereas inhibition of FAO with etomoxir did not alter SOX9 levels (Extended Data Fig. 7j). Moreover, knockdown of SOX9 in

growth plate-derived chondrocytes induced not only loss of typical chondrocyte characteristics such as cobblestone-like morphology and expression of *Col2a1* and *Acan* (Extended Data Fig. 7k, l), but also increased expression of *Cpt1a* and *Acadl* (Extended Data Fig. 7l) and the rate of FAO in chondrocytes (Fig. 3g, h). By contrast, overexpression of SOX9 in calvarial osteoblasts decreased FAO (Fig. 3h). SOX9 thus acts as a metabolic regulator in chondrogenic cells by suppressing FAO.

FOXOs induce SOX9 upon lipid starvation

We next examined how lipids regulate SOX9 levels. Transcriptomics showed robust upregulation of *Sox9* expression in C3H10T1/2 cells

starting 1 h after serum deprivation and increased expression of several other, but not all, chondrogenic markers from 3 h onwards (Extended Data Fig. 8a). Differential expression analysis showed that 678 (1 h), 4,022 (3 h) and 3,811 (6 h) genes were significantly upregulated by serum deprivation, including *Sox9* as one of the top hits at all time points (Fig. 4a, Extended Data Fig. 8b). A total of 757 (1 h), 2,167 (3 h) and 3,872 (6 h) genes were significantly downregulated, including genes associated with proliferation (*Egr3*, *Dusp5* and *Erff1*), skeletal stem cells (*Nes* and *Itga5*) and osteogenesis (*Spp1* and *Adam19*) (Fig. 4a; Extended Data Fig. 8b). Transcription factor-binding-motif analysis²⁶ of the top-100 overexpressed genes at each time point showed strong enrichment of the FOXO/forkhead motif (Fig. 4b, Extended Data Fig. 8c). We confirmed that serum deprivation increases nuclear FOXO1 and FOXO3a in C3H10T1/2 cells (Fig. 4c, d) and active FOXO levels in C3H10T1/2 and skeletal stem cells, an effect prevented by exogenous fatty acids (Fig. 4e, Extended Data Fig. 8d–f), indicating that extracellular lipids control FOXO activity. More specifically, FOXO1 and FOXO3a showed increased binding to the *Sox9* promoter during serum deprivation (Fig. 4f, Extended Data Fig. 8g), and the FOXO inhibitor AS1842856 prevented induction of SOX9 during lipid deprivation in all cell types (Fig. 4g, Extended Data Fig. 8h). Similar results were obtained using a CRISPR–Cas9 approach to conditionally delete *Foxo1* and *Foxo3a* (also known as *Foxo3*) in C3H10T1/2 cells, or using short hairpin RNAs (shRNAs) in skeletal stem cells (Extended Data Fig. 8i, j). These data demonstrate that FOXOs directly control *Sox9* transcription during lipid deprivation.

We next confirmed the relation between lipid deprivation, FOXOs and SOX9 during bone healing. First, the presence of the filter (0.2 µm) during bone-graft healing increased the number of cells positive for nuclear FOXO3a in the central periosteal region (Fig. 4h), similar to the increase in SOX9⁺ cells (Fig. 1e). Second, stimulation of fatty acid signalling using the FFAR1/4 agonist GW9508 during fracture healing strongly reduced the number of FOXO3a⁺ nuclei in the periosteal callus (Extended Data Fig. 8k), correlating with reduced amounts of cartilage (Fig. 2h). Third, skeletal stem cells with FOXO1 and FOXO3 inactivation failed to engraft into tibial fractures (Extended Data Fig. 8l), which may be owing to their inability to increase SOX9 levels upon lipid deprivation, or a general failure to survive transplantation-associated stress. Finally, local daily injection of the FOXO inhibitor AS1842856 during fracture healing reduced the amount of cartilage while not affecting new bone formation (Fig. 4i). Thus, FOXO signalling in vivo is negatively regulated by lipid availability and is required for skeletal progenitor cell chondrogenesis and survival during bone healing.

Discussion

On the basis of our findings, we propose a model in which the local vasculature, through supply of lipids, influences skeletal progenitor differentiation during fracture healing (Fig. 4j). Cells close to blood vessels become osteoblasts, which depend on FAO to support their metabolic demands. Skeletal progenitors in poorly vascularized regions sustain FAO for a short time by mobilizing intracellular lipid stores and then activate FOXO signalling as a result of exogenous lipid starvation. Nuclear localization of FOXOs promotes expression of SOX9, which induces chondrogenic commitment and suppresses FAO to allow long-term cell survival.

Low lipid levels are thus the main nutritional determinant for chondrogenic commitment of skeletal progenitor cells, rather than lack of oxygen or glucose^{19,20,27}, although growth factors are indispensable to activate the full chondrogenic-differentiation program^{1,2,9}. In contrast to osteoblasts^{28,29}, we find that chondrocytes are largely independent of FAO, consistent with poor diffusion of fatty acids in cartilage tissue. This metabolic independence from extracellular lipids would therefore be beneficial in the avascular cartilage environment. FAO in

chondrocytes is suppressed by SOX9, attributing a novel metabolic regulatory role to this transcription factor. Mechanistically, reduced lipid availability is translated into SOX9 production through FOXOs, well-known regulators of the cellular response to metabolic stress³⁰. We propose lipid starvation as an additional trigger for FOXO activation, although the full signalling cascade and exact lipid sensor remain unknown. Of interest, osteoarthritis is associated with increased angiogenesis and FAO^{31,32} but reduced SOX9 levels and FOXO activity^{33,34}. Our results show that all these phenomena may be connected to local lipid availability, suggesting that manipulation of lipid metabolism could be of therapeutic interest. More generally, our findings show that local nutrient levels can decide stem-cell lineage choice through direct transcriptional changes. As a consequence, the metabolic profile of a mature cell may reflect microenvironmental constraints as much as particular cellular needs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2050-1>.

- Kronenberg, H. M. Developmental regulation of the growth plate. *Nature* **423**, 332–336 (2003).
- Roberts, S. J., van Gestel, N., Carmeliet, G. & Luyten, F. P. Uncovering the periosteum for skeletal regeneration: the stem cell that lies beneath. *Bone* **70**, 10–18 (2015).
- Hallmann, R., Feinberg, R. N., Latker, C. H., Sasse, J. & Risau, W. Regression of blood vessels precedes cartilage differentiation during chick limb development. *Differentiation* **34**, 98–105 (1987).
- Yin, M. & Pacifici, M. Vascular regression is required for mesenchymal condensation and chondrogenesis in the developing limb. *Dev. Dyn.* **222**, 522–533 (2001).
- Maes, C. et al. Placental growth factor mediates mesenchymal cell development, cartilage turnover, and bone remodeling during fracture repair. *J. Clin. Invest.* **116**, 1230–1242 (2006).
- Taylor, D. K. et al. Thrombospondin-2 influences the proportion of cartilage and bone during fracture healing. *J. Bone Miner. Res.* **24**, 1043–1054 (2009).
- Miclaú, K. R. et al. Stimulating fracture healing in ischemic environments: does oxygen direct stem cell fate during fracture healing? *Front. Cell Dev. Biol.* **5**, 45 (2017).
- Tiyapatanaputi, P. et al. A novel murine segmental femoral graft model. *J. Orthop. Res.* **22**, 1254–1260 (2004).
- Stegen, S., van Gestel, N. & Carmeliet, G. Bringing new life to damaged bone: the importance of angiogenesis in bone repair and regeneration. *Bone* **70**, 19–27 (2015).
- Carlier, A. et al. MOSAIC: a multiscale model of osteogenesis and sprouting angiogenesis with lateral inhibition of endothelial cells. *PLOS Comput. Biol.* **8**, e1002724 (2012).
- Carlier, A., Geris, L., van Gestel, N., Carmeliet, G. & Van Oosterwyck, H. Oxygen as a critical determinant of bone fracture healing—a multiscale model. *J. Theor. Biol.* **365**, 247–264 (2015).
- Zhao, L., Li, G., Chan, K. M., Wang, Y. & Tang, P. F. Comparison of multipotent differentiation potentials of murine primary bone marrow stromal cells and mesenchymal stem cell line C3H10T1/2. *Calcif. Tissue Int.* **84**, 56–64 (2009).
- van Gestel, N. et al. Engineering vascularized bone: osteogenic and proangiogenic potential of murine periosteal cells. *Stem Cells* **30**, 2460–2471 (2012).
- Debnath, S. et al. Discovery of a periosteal stem cell mediating intramembranous bone formation. *Nature* **562**, 133–139 (2018).
- Duchamp de Lageneste, O. et al. Periosteum contains skeletal stem cells with high bone regenerative potential controlled by Periostin. *Nat. Commun.* **9**, 773 (2018).
- Chan, C. K. et al. Identification and specification of the mouse skeletal stem cell. *Cell* **160**, 285–298 (2015).
- Amarilio, R. et al. HIF1α regulation of Sox9 is necessary to maintain differentiation of hypoxic prechondrogenic cells during early skeletogenesis. *Development* **134**, 3917–3928 (2007).
- Robins, J. C. et al. Hypoxia induces chondrocyte-specific gene expression in mesenchymal cells in association with transcriptional activation of Sox9. *Bone* **37**, 313–322 (2005).
- Shapiro, I. M. & Srinivas, V. Metabolic consideration of epiphyseal growth: survival responses in a taxing environment. *Bone* **40**, 561–567 (2007).
- Stegen, S. et al. HIF-1α metabolically controls collagen synthesis and modification in chondrocytes. *Nature* **565**, 511–515 (2019).
- Baryawno, N. et al. A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. *Cell* **177**, 1915–1932.e16 (2019).
- Kim, C. et al. Studying arrhythmic right ventricular dysplasia with patient-specific iPSCs. *Nature* **494**, 105–110 (2013).
- Singh, R. et al. Autophagy regulates lipid metabolism. *Nature* **458**, 1131–1135 (2009).
- Rambold, A. S., Cohen, S. & Lippincott-Schwartz, J. Fatty acid trafficking in starved cells: regulation by lipid droplet lipolysis, autophagy, and mitochondrial fusion dynamics. *Dev. Cell* **32**, 678–692 (2015).

25. Tsukamoto, S. et al. Autophagy is essential for preimplantation development of mouse embryos. *Science* **321**, 117–120 (2008).
26. Imrichová, H., Hulselmans, G., Atak, Z. K., Potier, D. & Aerts, S. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res.* **43** (W1), W57–W64 (2015).
27. Shang, J., Liu, H., Li, J. & Zhou, Y. Roles of hypoxia during the chondrogenic differentiation of mesenchymal stem cells. *Curr. Stem Cell Res. Ther.* **9**, 141–147 (2014).
28. Frey, J. L. et al. Wnt–Lrp5 signaling regulates fatty acid metabolism in the osteoblast. *Mol. Cell. Biol.* **35**, 1979–1991 (2015).
29. Kim, S. P. et al. Fatty acid oxidation by the osteoblast is required for normal bone acquisition in a sex- and diet-dependent manner. *JCI Insight* **2**, 92704 (2017).
30. Eijkelenboom, A. & Burgering, B. M. FOXOs: signalling integrators for homeostasis maintenance. *Nat. Rev. Mol. Cell Biol.* **14**, 83–97 (2013).
31. Ashraf, S. & Walsh, D. A. Angiogenesis in osteoarthritis. *Curr. Opin. Rheumatol.* **20**, 573–580 (2008).
32. Ratneswaran, A. et al. Peroxisome proliferator-activated receptor δ promotes the progression of posttraumatic osteoarthritis in a mouse model. *Arthritis Rheumatol.* **67**, 454–464 (2015).
33. Zhong, L., Huang, X., Karperien, M. & Post, J. N. Correlation between gene expression and osteoarthritis progression in human. *Int. J. Mol. Sci.* **17**, E1126 (2016).
34. Akasaki, Y. et al. Dysregulated FOXO transcription factors in articular cartilage in aging and osteoarthritis. *Osteoarthritis Cartilage* **22**, 162–170 (2014).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized, unless otherwise mentioned. The investigators were not blinded to allocation during experiments and outcome assessment.

Mice

C57BL/6J mice, 129/Sv mice (Janvier Labs), B6.Cg-Tg(CAG-eGFP) mice³⁵, B6.Cg-Tg(Col1a1-cre/ERT2,-DsRed)1Smkm/J mice³⁶, B6;129S4-Sox9tm1.1Tlu/J mice and B6.Cg-Tg(CAG-DsRed*MST)1Nagy/J mice (The Jackson Laboratory) were used in this study. Unless otherwise specified, both male and female mice were used for all experiments. All animal experiments were conducted according to the regulations and with approval of the Animal Ethics Committee of the KU Leuven.

Mouse bone-healing models

The femoral segmental bone-graft model was adapted from a previously described model⁸. Eight- to ten-week old male C57BL/6J mice were anaesthetized with a ketamine–xylazine mixture (100 mg per kg ketamine and 15 mg per kg xylazine) and the right femur was exposed. A mid-diaphyseal 4-mm bone segment was excised with a 6.5-mm diamond saw disk (Codema), briefly washed in saline to remove the bone marrow (periosteum not removed) and the segment was subsequently reimplanted in the defect (autograft). To investigate the contribution of donor cells, grafts were isolated from CAG-eGFP mice (periosteum not removed) and transplanted in wild-type littermates. To obtain devitalized allografts, 4-mm bone segments were isolated from 129/Sv mice, washed in saline to remove the bone marrow, scraped to remove the periosteum, sterilized in 70% ethanol and frozen at -80°C for at least 1 week. After graft implantation, the defect was stabilized with an intramedullary metal pin (22 gauge needle). To create a compromised host environment, a polycarbonate filter with a pore size of $30\text{ }\mu\text{m}$ or $0.2\text{ }\mu\text{m}$ (Sterilitech) was inserted between the muscle and the graft at the time of surgery.

The tibial fracture healing model was performed as previously described⁵. For studies with the FFAR1/4 agonist GW9508 mice were treated daily by subcutaneously injecting $50\text{ }\mu\text{l}$ of a $200\text{ }\mu\text{M}$ GW9508 (Cayman Chemical) solution or vehicle (0.2% DMSO in saline) at the fracture site. For fatty acid delivery, mice were treated daily by subcutaneously injecting $20\text{ }\mu\text{l}$ corn oil (Sigma) or control solution (saline) at the fracture site. For studies with the FOXO inhibitor AS1842856, mice were treated daily by subcutaneously injecting $50\text{ }\mu\text{l}$ of a $10\text{ }\mu\text{M}$ AS1842856 (Calbiochem) solution or vehicle (0.1% DMSO in saline) at the fracture site. For metabolite labelling experiments mice were injected intravenously with the fluorescent fatty acid analogue BODIPY 558/568 C12 (Red-C12; Invitrogen) at $1\text{ mg per kg body weight}$ and the fluorescent glucose analogue 2-(*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl) amino)-2-deoxyglucose (2-NBDG; Invitrogen) at $12.5\text{ mg per kg body weight}$, 15 min before euthanasia. For skeletal stem cell transplantations, $100,000$ cells (sh*Cpt1a* experiments) or $20,000$ cells (sh*Foxo1* and sh*Foxo3a* experiments) were resuspended in $5\text{ }\mu\text{l}$ of a 5 mg ml^{-1} collagen gel (rat tail collagen type I, Corning) and transplanted at the fracture site at the time of surgery.

Micro-computed tomography analysis

Mice were euthanised at 2 or 4 weeks after surgery and grafted bones were isolated. For bone analysis, samples were scanned using the high resolution SkyScan 1172 micro-computed tomography (microCT) system (Bruker-microCT) at a pixel size of $10\text{ }\mu\text{m}$ with 50 kV tube voltage and 0.5 mm aluminium filter. To reduce the metal artefacts induced by the presence of the intramedullary pin, microCT projection data was reconstructed using an iterative reconstruction technique and projection completion³⁷. Custom software was made in MeVisLab (MeVis Medical Solutions) to visualize and analyse the obtained microCT images.

The boundary between graft and callus was manually delineated and mineralized tissue was segmented using hysteresis thresholding. For visual representation grafts are represented in a different colour to callus and host bone. The coverage ratio was calculated as the percentage of the graft surface that is covered with callus by determining whether the normal line to the graft surface encounters mineralized callus, for each point of the graft surface.

For visualization and quantification of the vasculature, mice were anaesthetized with a ketamine–xylazine–heparin mixture (100 mg per kg ketamine, 15 mg per kg xylazine and $1,000\text{ U per kg heparin}$) and successively perfused with 10 ml of heparinized saline (100 U ml^{-1}), 10 ml of a 10% neutral-buffered formalin solution, 10 ml of saline and 5 ml of a preheated 30% barium sulfate solution (Micropaque, Guerbet) containing 2% gelatine. After perfusion, animals were placed on ice for at least 1 h and subsequently kept at 4°C overnight to allow the gelatine to solidify, before removing the grafted hindlimbs for dual-energy microCT analysis^{38,39}. Two microCT scans of each sample were taken on the SkyScan 1172 microCT system with effective beam energy below (50 kV tube voltage with 0.5 mm aluminium filter) and above (100 kV tube voltage with 0.5 mm aluminium and 0.038 mm copper filter) the K-edge energy of barium sulfate, both with an image pixel size of $5\text{ }\mu\text{m}$. By combining the low and high energy acquisitions, an image of the (barium sulfate-perfused) vasculature only was reconstructed as described^{38,39} and a segmentation of the vasculature was obtained by thresholding this image. A segmentation of the bone was obtained by thresholding the bone and vasculature out of the low energy reconstruction and removing the calculated vasculature from it. After delineating a $250\text{-}\mu\text{m}$ -wide region of interest around the graft surface using a custom made MeVisLab software package, calculation of the number of blood vessels and the average vessel thickness was performed using the CTAn software (Bruker-microCT).

Immunohistochemistry

To isolate bones for histological analysis, mice were anaesthetized with ketamine–xylazine–heparin and perfused with 10 ml of heparinized saline followed by 10 ml of 2% paraformaldehyde in PBS. Isolated bones were further fixed in 2% paraformaldehyde overnight and decalcified in EDTA for 14 days at 4°C . Samples were either embedded in paraffin and sectioned at $4\text{ }\mu\text{m}$, embedded in agarose for vibratome sections ($100\text{ }\mu\text{m}$ thick) or embedded in NEG-50 frozen section medium (Richard-Allen Scientific) and sectioned at $7\text{ }\mu\text{m}$ using the CryoJane Tape-Transfer System (Leica) for samples containing fluorescent protein-expressing cells. Staining with haematoxylin and eosin (H&E) and safranin O, terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) staining and immunohistochemical staining for BrdU, CD31 and COL2 are routinely used in our laboratory and have all been described previously^{5,13,40–43}. For SOX9, COL1, CPT1a, GLUT1 and FOXO3a immunohistochemical staining, sections were deparaffinised and blocked for 30 min in 0.1 M Tris-HCl, 0.15 M NaCl, pH 7.6 (TNT) with 0.5% Blocking Reagent (NEN, PerkinElmer) and 20% normal goat serum (DAKO). Subsequently, sections were incubated overnight with a rabbit anti-SOX9 primary antibody (Novus Biologicals; NBPI-85551; 1:100), rabbit anti-COL1 primary antibody (Novus Biologicals; NB600-408; 1:100), rabbit anti-CPT1a primary antibody (Cell Signaling Technology; no. 12252; 1:50), rabbit anti-GLUT1 primary antibody (Cell Signaling Technology; no. 12939; 1:100) or rabbit anti-FOXO3a primary antibody (Cell Signaling Technology, no. 2497; 1:100) diluted in TNT with 0.5% blocking reagent, followed by three washes with TNT containing 0.05% Tween-20. Next, slides were incubated for 1 h with an Alexa Fluor 546- or Alexa Fluor 488-conjugated goat anti-rabbit secondary antibody (Invitrogen; A-11010 and A-11034) diluted 1:200 in TNT/0.5% blocking reagent, washed and counterstained with Hoechst 33342 ($20\text{ }\mu\text{g ml}^{-1}$ in PBS; Invitrogen). Stainings omitting the primary antibody were used as negative controls.

Images were taken on a Zeiss Axioplan 2 light microscope, Zeiss LSM510-META NLO multi-photon confocal microscope or Zeiss LSM880 confocal laser scanning microscope. Histomorphometry was performed using the Zeiss AxioVision software, ImageJ software (National Institutes of Health) and CellProfiler software⁴⁴. Quantification of blood vessels or proliferating cells was performed by respectively counting CD31⁺ vessels or BrdU⁺ cells in a 250- μ m-wide region of interest adjacent to the graft surface. Apoptotic or chondrogenic cells were quantified by respectively counting the number of TUNEL⁺ or SOX9⁺ cells and the total number of cells in a 0.015 mm² region of interest near the graft surface at the centre of the graft. Quantification of cartilage was performed by outlining COL2⁺ or safranin O⁺ areas within the total callus area (for fractures and grafts) or the central graft callus area (half of total graft length). Quantification of woven bone was performed by outlining areas of macroscopically-defined immature bone within the total callus area. Quantification of FOXO3a⁺ nuclei was performed using the 'cell/particle counting and scoring' pipeline in CellProfiler, in a region of interest encompassing the total callus area (for fractures) or the central graft callus area (half of total graft length). For all quantifications, measurements were made on at least three different sections throughout the sample.

Computational model of bone-graft healing

We used a previously established multiscale computational framework of bone regeneration that quantitatively describes the interplay between cells, growth factors, nutrient levels and blood vessels^{10,11}. In short, this multiscale model combines ten partial differential equations of the taxis-reaction-diffusion type at the tissue level with a discrete agent-based approach at the vascular level, including eight intracellular variables for the endothelial cells. At the tissue level, the model accounts for the various key processes of intramembranous and endochondral ossification that occur during the soft and hard callus phase of bone healing. The partial differential equations describe the evolution in time and space of the skeletal progenitor cell density, fibroblast density, chondrocyte density, osteoblast density, fibrous matrix density, cartilaginous matrix density, bone matrix density, osteochondrogenic growth factor concentration, vascular growth factor concentration and nutrient concentration. For simplification purposes, only one generic osteochondrogenic growth factor and one nutrient parameter is included in the computational model, which respectively represent the effects of multiple growth factors (for example, transforming growth factors or bone morphogenetic proteins) and nutrients (such as oxygen, glucose, amino acids or lipids) present during bone healing. The assumption is made that the net result of all growth factors present will be to promote chondrogenesis and osteogenesis, and thus if local levels of the osteochondrogenic growth factor reach a certain threshold (modelled using a sixth-order Hill function) it will induce differentiation of skeletal progenitor cells. The decision on whether the end result of this differentiation event is chondrogenic or osteogenic is made by the nutrient parameter. The influence of the generic osteochondrogenic growth factor on skeletal progenitor cell differentiation is promoting chondrogenic differentiation when local nutrient levels are low, and promoting osteogenic differentiation when local nutrient levels are high. Cell types that are considered at the tissue scale (skeletal progenitor cells, chondrocytes, osteoblasts, fibroblasts) can migrate (only skeletal progenitor cells and fibroblasts), proliferate, differentiate and produce growth factors (generic osteochondrogenic growth factor or angiogenic growth factor) and extracellular matrix (cartilage, bone or fibrous tissue). Blood vessels are modelled at both a cellular level (representing the developing vasculature with discrete endothelial cells) and an intracellular level (that defines the internal dynamics of every endothelial cell), and serve as the nutrient source. At the cellular level, the development of the discrete vascular tree (composed of endothelial cells) is determined by three different processes, that is, sprouting (the formation of a new branch, headed

by a tip endothelial cell), vascular growth (the extension of the branch due to tip cell migration) and anastomosis (the fusion of two branches). An anastomosis between blood vessels allows for blood flow and the delivery of nutrients. The intracellular level considers a number of molecular players that govern endothelial cell movement (VEGFR2, DLL4, Notch and actin).

While the blood vessels are modelled discretely, continuous variables are used for nutrient density, bone density, cartilage density and fibrous tissue density (included in the model but not relevant for the current setup and therefore not shown). The colour scale for nutrients, bone and cartilage thus indicates a continuous gradient going from complete absence of a parameter ('0' value; nutrients, bone or cartilage are not present at that location) to complete saturation of a parameter ('1' value; a location is completely filled with nutrients, bone or cartilage). All values in between 0 and 1 represent partial filling of a location with a parameter. For the 'tissue' continuous variables (bone, cartilage, fibrous tissue), the sum of all tissues is 1, meaning that if a location is completely filled with bone (value '1'), no cartilage can exist at the same location (value '0'). However, since the variables are continuous, a specific location can contain both a fraction of bone and a fraction of cartilage. Tissues, nutrients and blood vessels are modelled in separate spaces and can thus 'co-exist' in the same location. Since the nutrient parameter is also continuous, it has an independent scale going from no nutrients (value '0') to saturating levels of nutrients (value '1', which we define as the level of nutrients found inside a modelled blood vessel).

By adapting the geometry and boundary conditions to the bone-graft setup, the influence of a filter placed in between graft and muscle on the healing process can be predicted *in silico*. Detailed information on the equations, parameter values and implementation can be found in ref.¹¹. Information on the boundary and initial conditions used in this study can be found in Extended Data Fig. 3.

Isolation of primary cells

Periosteal cells and trabecular osteoblasts were isolated from the long bones of 8–10-week-old mice as described¹³. For the isolation of periosteal cells, femurs and tibias were dissected free of muscle and connective tissue under sterile conditions. Subsequently, the epiphyses were protected from digestion by submerging them in 5% low melting point agarose (SeaPlaque, Lonza) and periosteal cells were isolated by enzymatic digestion using 3 mg ml⁻¹ collagenase II (Gibco) and 4 mg ml⁻¹ dispase (Gibco) in α -minimal essential medium (α -MEM; Gibco) supplemented with 1% penicillin/streptomycin (100 units ml⁻¹ and 100 μ g ml⁻¹ respectively; Gibco). Cells from the first digest (10 min) were discarded as they contain cells from remaining muscle and connective tissue, and periosteal cells were obtained by a subsequent 1 h digest. The cells were passed through a 70- μ m nylon mesh (BD Falcon), washed twice and cultured in α -MEM with 1% penicillin/streptomycin and 10% FBS (HyClone) in a humidified incubator at 37 °C with 5% CO₂. For the isolation of trabecular osteoblasts, femurs and tibias were cleaned thoroughly to remove muscle, connective tissue and periosteum. Subsequently, bones were incubated in collagenase–dispase (3 mg ml⁻¹ collagenase II and 4 mg ml⁻¹ dispase in α -MEM with 1% penicillin/streptomycin) for 20 min to remove remaining periosteal cells. Next, epiphyses were cut away, bone marrow was flushed out and the bone was cut into small pieces. Trabecular osteoblasts were isolated by incubating the bone fragments with collagenase–dispase for 30 min. Cells were passed through a 70- μ m nylon mesh, washed twice and cultured in α -MEM supplemented with 1% penicillin/streptomycin and 10% FBS at 37 °C with 5% CO₂. Cells from passage 2–3 were used for all experiments.

Growth plate-derived chondrocytes and calvarial osteoblasts were isolated from 3–5-day-old mice as described^{13,20}. For murine growth plate-derived chondrocytes the resting zones of the growth plates from the distal femora and proximal tibiae were dissected free from surrounding tissue and pre-digested for 30 min with 1 mg ml⁻¹ collagenase II in α -MEM with 1% penicillin/streptomycin on a shaker at

Article

room temperature. Cartilage fragments were then washed twice and subsequently digested for 3 h in a 2 mg ml⁻¹ collagenase II solution in α -MEM with 1% penicillin/streptomycin on a shaker at 37 °C. The cell suspension was then filtered through a 40- μ m nylon mesh, washed and cultured in α -MEM supplemented with 1% penicillin/streptomycin and 10% FBS at 37 °C with 5% CO₂. Calvarial osteoblasts were prepared by 6 sequential 15-min digestions of calvaria from 3–5-day-old mice in PBS containing 1 mg ml⁻¹ collagenase II and 2 mg ml⁻¹ dispase. Cells isolated in fractions 2–6 were pooled and cultured in α -MEM supplemented with 1% penicillin/streptomycin and 10% FBS at 37 °C with 5% CO₂. Cells from passage 2–3 were used for all experiments.

For isolation of rib chondrocytes, anterior rib cages were dissected from 5-day-old mice. Isolated rib cages were pre-digested on a shaker for 30 min at room temperature with 1 mg ml⁻¹ collagenase II (Gibco) dissolved in α -MEM supplemented with 1% penicillin/streptomycin. Rib fragments were subsequently digested for 3 h in a 2 mg ml⁻¹ collagenase II solution in α -MEM with 1% penicillin/streptomycin on a shaker at 37 °C. The obtained cell suspension of the second digest was filtered through a 40- μ m nylon mesh and single cells were recovered by centrifugation. Cells were cultured in a humidified incubator at 37 °C with 5% CO₂ in α -MEM supplemented with 1% penicillin/streptomycin and 10% FBS. Cells from passage 2–3 were used for all experiments.

Isolation of mouse skeletal stem cells was adapted from a previously described protocol¹⁶. Long bones of 3–5-day-old mice were dissected, muscle was cleared away carefully to preserve the periosteum and bones were minced using a scalpel. Bone fragments were then digested in α -MEM supplemented with 3 mg ml⁻¹ collagenase II, 4 mg ml⁻¹ dispase (both from Gibco) and 100 U ml⁻¹ DNase I (Sigma) at 37 °C for 3 sequential 15-min digests. Cell fractions were pooled and passed through a 70- μ m nylon mesh, washed with PBS containing 2% FBS and stained with antibodies against CD45, TER119, TIE2, CD105, CD90.2, CD249 (also known as 6C3) (BioLegend), CD51 (BD Pharmingen) and CD200 (eBioscience), and with the phylophen dye 7-aminoactinomycin D (7AAD; BD Pharmingen). Immunophenotypically-defined skeletal stem cells¹⁶ (7AAD⁻CD45⁻TER119⁻TIE2⁻CD51⁺CD105⁻CD90.2⁻CD249⁻CD200⁺; Extended Data Fig. 9a) were sorted on a BD FACS Aria II (BD Biosciences). Single colour controls were used to set compensations and fluorescence minus one controls were used to set gates. Sorted cells were cultured in a humidified incubator at 37 °C with 2% O₂ and 7.5% CO₂ in α -MEM supplemented with 1% penicillin/streptomycin and 10% FBS. For metabolic analyses, skeletal stem cells were grown in atmospheric O₂ levels with 5% CO₂ to enable direct comparison with other cell types. Cells from passage 2–3 were used for all experiments. For flow cytometric analysis of culture-expanded skeletal stem cells, cells were gated again for the CD51⁺CD105⁻CD90.2⁻CD249⁻CD200⁺ population to limit analysis to the stem cell fraction.

For the isolation of skeletal muscle-derived cell populations, hindlimb skeletal muscles, including quadriceps, soleus, gastrocnemius and tibialis anterior, were dissected from 8-week-old CAG-DsRed mice, minced using a scalpel and digested in α -MEM medium supplemented with 3 mg ml⁻¹ collagenase II, 4 mg ml⁻¹ dispase and 100 U ml⁻¹ DNase I at 37 °C for 60 min. Every 15 min, samples were pipetted up and down vigorously using a 10-ml serological pipette to break up tissue fragments. Cell suspensions were passed through a 70- μ m nylon mesh, washed with PBS containing 2% FBS and stained with antibodies against CD45, TER119, CD31, F4/80 and CD146 (BioLegend), and with 7AAD (BD Pharmingen). Immunophenotypically-defined macrophages (7AAD⁻CD45⁺F4/80⁺), endothelial cells (7AAD⁻CD45⁻TER119⁻F4/80⁻CD31⁺CD146⁺) and pericytes (7AAD⁻CD45⁻TER119⁻F4/80⁻CD31⁻CD146⁺) (Extended Data Fig. 9b) were sorted on a BD FACS Aria II. Single colour controls were used to set compensations and fluorescence minus one controls were used to set gates. Sorted cells were used for co-cultures with periosteal cells in micromasses.

Cell lines

The C3H10T1/2 cell line, used as a skeletal progenitor cell model¹², was obtained from the RIKEN Cell Bank and cultured in a humidified incubator at 37 °C with 5% CO₂ in α -MEM with 1% penicillin/streptomycin and 10% FBS. Cells were routinely tested and found negative for mycoplasma contamination.

Nutrient-deprivation assays

Cells were seeded at 3,000 cells per cm² in basal DMEM (glucose- and glutamine-free; Gibco) supplemented with 1% penicillin/streptomycin, 5 mM D-(+)-glucose (Sigma-Aldrich), 2 mM L-glutamine (Gibco), 1 mM sodium pyruvate (Gibco) and 10% dialysed FBS (HyClone). After 24 h, cells were washed with PBS and switched to control medium (basal DMEM with 1% penicillin/streptomycin, 5 mM glucose, 2 mM L-glutamine, 1 mM sodium pyruvate and 10% dialysed FBS), SD medium (basal DMEM with 1% penicillin/streptomycin, 5 mM glucose, 2 mM L-glutamine, 1 mM sodium pyruvate and 1% dialysed FBS), glucose-deprivation medium (basal DMEM with 1% penicillin/streptomycin, 0.5 mM glucose, 2 mM L-glutamine, 1 mM sodium pyruvate and 10% dialysed FBS), glutamine-deprivation medium (basal DMEM with 1% penicillin/streptomycin, 5 mM glucose, 0.2 mM L-glutamine, 1 mM sodium pyruvate and 10% dialysed FBS), CND medium (basal DMEM with 1% penicillin/streptomycin, 0.5 mM glucose, 0.2 mM L-glutamine, 1 mM sodium pyruvate and 1% dialysed FBS) or LRS medium (basal DMEM with 1% penicillin/streptomycin, 5 mM glucose, 2 mM L-glutamine, 1 mM sodium pyruvate and 10% lipid-reduced FBS). LRS was made by mixing FBS with fumed silica (Sigma) at 20 mg ml⁻¹ for 3 h at room temperature, followed by centrifugation at 2,000g for 15 min and filtration of the supernatant through a 0.45- μ m-pore-size filter.

In certain experiments cultures were supplied with actinomycin D (transcription inhibitor; Sigma-Aldrich), cycloheximide (translation inhibitor; Sigma-Aldrich), chloroquine (lysosomal inhibitor; Sigma-Aldrich) or etomoxir (CPT1 inhibitor; Merck-Millipore) at the concentrations indicated in the text. For lipid rescue experiments, SD medium was supplemented with very low density lipoproteins (VLDL; Calbiochem) at a concentration of 607 μ g triglycerides per ml FBS, palmitic or oleic acid (Sigma-Aldrich) at the indicated concentrations or a mixture of PUFAs (10 μ M linoleic acid, 15 μ M α -linolenic acid, 10 μ M arachidonic acid and 15 μ M docosahexaenoic acid; all from Sigma-Aldrich). Triglycerides were incubated in FBS for 30 min at 37 °C and fatty acids (dissolved in ethanol) were complexed to fatty acid-free bovine serum albumin (BSA) (Sigma-Aldrich) for 1 h at 37 °C before adding to the culture medium, as described previously⁴⁵. All supplements were added at the start of the experiment and were present for the entire duration of the cultures.

Differentiation assays

To assess chondrogenic differentiation, 150,000 periosteal cells were resuspended in 10 μ l of control medium and seeded as micromasses in the middle of a 24-well plate. Cells were allowed to attach for 1 h at 37 °C, after which 0.5 ml of control, SD or LRS medium containing 10 ng ml⁻¹ recombinant human TGF β 1 (Peprotech), 50 μ M L-ascorbic acid 2-sulfate (Sigma-Aldrich) and 20 μ M Y-27632 (Rho kinase inhibitor; Axon Medchem)⁴⁶ was added to the wells. Medium was refreshed every other day and after 9 days micromasses were either stained with alcian blue or used for RNA isolation. For chondrogenic differentiation in the presence of muscle-derived cells micromasses were made using 100,000 periosteal cells derived from Sox9-GFP mice and 50,000 skeletal muscle-derived macrophages, endothelial cells, pericytes or unsorted cells obtained from CAG-DsRed mice.

For chondrogenic differentiation in pellets 200,000 periosteal cells were placed in a 5-ml polystyrene tube in 1 ml of control, SD or LRS medium containing 10 ng ml⁻¹ recombinant human TGF β 1 (Peprotech) and 50 μ M L-ascorbic acid 2-sulfate (Sigma-Aldrich), supplemented with

vehicle (1% ethanol in 4% fatty acid-free BSA in saline), 60 μ M oleate or a mixture of PUFA (10 μ M linoleic acid, 15 μ M α -linolenic acid, 10 μ M arachidonic acid and 15 μ M docosahexaenoic acid) complexed to fatty acid-free BSA. Tubes were centrifuged for 5 min at 500g and placed in a humidified incubator at 37 °C. Medium was changed every 3 days and after 21 days pellets were fixed in 4% paraformaldehyde for 10 min and processed for paraffin histological sectioning.

For osteogenic differentiation, periosteal cells were seeded cells at 30,000 cells per cm^2 in control medium and cultured for 3 days in order to reach full confluence. Cells were then switched to control, SD or LRS medium containing 50 μ M L-ascorbic acid 2-sulfate and 10 mM β -glycerophosphate (Sigma-Aldrich). After 21 days, cells were either stained with alizarin red S to detect mineralization or used for RNA isolation.

Metatarsal cultures

Metatarsal rudiments were dissected from E16.5 Col1a1-cre/ERT2, -DsRed embryos and stripped of skin. The middle three metatarsals were kept together as triads and cultured for 7 days on a Falcon insert membrane (pore size 0.4 μ m) in 12-well plates in 1 ml of BGJb culture medium (Gibco) supplemented with 25 $\mu\text{g ml}^{-1}$ L-ascorbic acid 2-sulfate, 10 mM β -glycerophosphate, and FBS (10% or 1%)⁴⁰. When indicated a mixture of PUFAs (10 μ M linoleic acid, 15 μ M α -linolenic acid, 10 μ M arachidonic acid and 15 μ M docosahexaenoic acid) complexed to fatty acid-free BSA or vehicle (1% ethanol in 4% fatty acid-free BSA in saline) was added to the culture medium. At the end of the cultures the metatarsals were fixed overnight in 2% paraformaldehyde in PBS and processed for histochemistry or immunohistochemistry.

Flow cytometry

Cell death was detected using annexin V-FITC and propidium iodide (Dead Cell Apoptosis Kit; Invitrogen), or using active caspase 3-FITC (FITC Active Caspase-3 Apoptosis Kit; BD Pharmingen). Proliferation was assessed by staining with a PE-conjugated mouse anti-Ki-67 antibody (BD Pharmingen; #556027; 1:10) and Hoechst 33342 (40 $\mu\text{g ml}^{-1}$; Invitrogen) after fixation and permeabilization of the cells (BD Cytofix/Cytoperm Kit, BD Biosciences). Intracellular SOX9 levels were quantified by staining with an Alexa Fluor 647-conjugated rabbit anti-SOX9 antibody (Cell Signaling Technology; no. 71273; 1:100) after fixation and permeabilization of the cells. Gating for SOX9^{high} cells was set to have approximately 10% SOX9^{high} cells in control conditions. Single colour controls were used to set compensations and fluorescence minus one controls were used to set gates.

Immunocytochemistry

For immunofluorescence microscopy, cells grown on coverslips were fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton-X100 in PBS and blocked with PBS containing 5% BSA, 5% normal goat serum and 0.5% Tween-20. Next, cells were incubated overnight at 4 °C with primary antibodies (rabbit anti-FOXO1, Cell Signaling Technology, no. 2880, 1:100; rabbit anti-FOXO3a, Cell Signaling Technology, no. 2497, 1:100) in blocking buffer, followed by three washes with PBS/Tween-20. Slides were subsequently incubated for 2 h with secondary antibodies (Alexa Fluor 488-conjugated goat anti-rabbit; 1:500) in PBS containing 5% BSA and 0.5% Tween-20, washed and counterstained with Hoechst 33342. Stainings omitting the primary antibody were used as negative controls.

For staining of lipid droplets with 1,6-diphenyl-1,3,5-hexatriene (DPH), cells grown on coverslips were washed with PBS and fixed with 3.7% formaldehyde in PBS. DPH staining solution was prepared by diluting a 2 mM DPH (Sigma-Aldrich) stock (in DMSO) in PBS to a final concentration of 4 μ M as previously described⁴⁷. Cells were stained with DPH for 30 min, washed and nuclei were counterstained using TO-PRO-3 (Molecular Probes).

For tracking lipid movement between lipid droplets and mitochondria, cells were incubated with the fluorescent fatty acid analogue

BODIPY 558/568 C12 (Red-C12; Invitrogen) at 1 μ M in culture medium for 16 h (ref.²⁴). Cells were then washed three times with culture medium, incubated for 1 h in culture medium to allow the fluorescent lipids to incorporate into lipid droplets, and then chased for the time indicated in control or SD medium. Mitochondria were labelled with 100 nM MitoTracker Deep Red FM (Invitrogen) for 30 min before the end of the experiment. Cells were fixed and lipid droplets were stained with DPH as described above.

For measurement of autophagic flux, cells grown on coverslips were transfected with 1 μ g of an RFP-GFP-LC3 tandem construct⁴⁸ using the X-tremeGENE HP transfection reagent (Roche) according to the manufacturer's instructions. After 24 h, cells were washed with PBS and used for subsequent experiments. Since the GFP-LC3 loses fluorescence owing to lysosomal acidic and degradative conditions but the RFP-LC3 does not, autophagosomes in the cell are seen as green-yellow puncta, whereas autophagolysosomes are red.

Images were taken on a Zeiss LSM10-META NLO multi-photon confocal microscope or Zeiss LSM880 confocal laser scanning microscope, and prepared using Adobe Photoshop CS5 (Adobe) and ImageJ. LC3 puncta and DPH⁺ lipid droplets per cell were counted manually in ImageJ, while overlap between MitoTracker and Red-C12 in manually delineated cells was performed using the 'co-localization' plugin for ImageJ after thresholding of individual frames.

Western blot analysis

Total cell lysates were obtained by lysing cells in 25 mM Tris-HCl buffer (pH 7.6) containing 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, 1 \times cOmplete protease inhibitor cocktail (Roche) and 1 \times PhosSTOP phosphatase inhibitor cocktail (Roche). For cytoplasmic and nuclear extracts, cells were first lysed in 20 mM Hepes (pH 7.9) containing 10 mM KCl, 1.5 mM MgCl₂, 1 mM EDTA, 0.5% NP40, 1 mM DTT, 1 mM Na₃VO₄, 20 mM NaF, 1 mM PMSEF, 5 $\mu\text{g ml}^{-1}$ aprotinin, 5 $\mu\text{g ml}^{-1}$ leupeptin and 0.33 $\mu\text{g ml}^{-1}$ antipain. Following 15 min incubation at 4 °C, the cell lysates were passed 10 times through a 26 gauge needle. After centrifugation for 1 min at 18,000g, the supernatant (cytoplasmic proteins) was removed and the pellet containing the nuclear protein fraction was resuspended in 50 mM Hepes (pH 7.9) containing 500 mM NaCl, 1% NP40, 5 $\mu\text{g ml}^{-1}$ aprotinin, 5 $\mu\text{g ml}^{-1}$ leupeptin and 0.33 $\mu\text{g ml}^{-1}$ antipain, and sonicated. Proteins (10 μ g, except for detection of LC3 for which 20 μ g was used) were separated by SDS-PAGE and transferred to a nitrocellulose membrane (GE Healthcare). Membranes were blocked with 5% dry milk in Tris-buffered saline with 0.1% Tween-20 for 30 min at room temperature and incubated overnight at 4 °C with primary antibodies (rabbit anti-SOX9, Novus Biologicals, NBP1-85551, 1:2,000; rabbit anti-FOXO1, Cell Signaling Technology, no. 2880, 1:1,000; rabbit anti-FOXO3a, Cell Signaling Technology, no. 2497, 1:1,000; rabbit anti-LC3B, Cell Signaling Technology, no. 3868, 1:500; mouse anti- β -actin, Sigma, A5441, 1:10,000; mouse anti-lamin A/C, Santa Cruz Biotechnology, sc-376248, 1:5,000) diluted in blocking buffer. Signals were detected by enhanced chemiluminescence (Perkin Elmer) after incubation with HRP-conjugated secondary antibodies (DAKO). For gel source data, see Supplementary Fig. 1.

Metabolic assays

Glucose and lactate levels in culture medium were measured on a AU640 Chemistry Analyzer (Beckman Coulter). Glucose consumption was calculated by subtracting the remaining amount of glucose in the culture medium after 24 h of incubation with cells from the amount of glucose in unspent medium, and normalized for time and for cell number via DNA quantification. In a similar way, lactate secretion was calculated by subtracting lactate levels in unspent medium from the levels in medium incubated for 24 h with cells. Oxygen consumption was determined on a Seahorse XF24 Analyzer (Seahorse Bioscience) using 50,000 cells per well. The assay medium was unbuffered DMEM (Sigma) supplemented with 5 mM D-glucose and 2 mM L-glutamine,

Article

pH 7.4. For quantification of FAO-linked oxygen consumption the difference in OCR before and after injection of etomoxir (100 μM final concentration) was calculated²².

For measurement of glycolysis, cells were incubated for 6 h in growth medium containing 0.3 $\mu\text{Ci ml}^{-1}$ [$5\text{-}^3\text{H}$]D-glucose (PerkinElmer). The culture medium was then transferred into glass vials sealed with rubber caps. $^3\text{H}_2\text{O}$ was captured in hanging wells containing a Whatman paper soaked with H_2O over a period of 48 h at 37 °C to reach saturation⁴⁹. Radioactivity was determined in the paper by liquid scintillation counting and values were normalized to DNA content.

For glucose oxidation, cells were incubated for 6 h in growth medium containing 0.6 $\mu\text{Ci ml}^{-1}$ [$6\text{-}^{14}\text{C}$]D-glucose (PerkinElmer). To stop cellular metabolism, 250 μl of a 2 M perchloric acid solution was added and wells were covered with a Whatman paper soaked with 1x hyamine hydroxide. $^{14}\text{CO}_2$ released during the oxidation of glucose was absorbed into the paper overnight at room temperature. Radioactivity in the paper was determined by liquid scintillation counting, and values were normalized to DNA content⁴⁹.

FAO was measured after incubation of the cells with 3 $\mu\text{Ci ml}^{-1}$ [$9,10\text{-}^3\text{H}$]palmitate (PerkinElmer), complexed to BSA, for 2 h. Then, the culture medium was transferred into glass vials sealed with rubber caps. $^3\text{H}_2\text{O}$ was captured in hanging wells containing a Whatman paper soaked with H_2O over a period of 48 h at 37 °C. Radioactivity in the paper was determined by liquid scintillation counting, and values were normalized to DNA content⁴⁹.

Metabolite diffusion assay

Diffusion rates were measured in custom-designed diffusion chambers according to a previously established protocol⁵⁰. Chambers were fabricated in a polydimethylsiloxane (PDMS) device on a glass substrate with medium reservoirs that contained fluorescent tracer molecules. 2-NBDG (342 Da) and BODIPY FL C16 (FL-C16; Invitrogen) complexed to fatty acid-free BSA (66.5 kDa) were used as fluorescent analogues to evaluate the diffusion rates of glucose and fatty acids, respectively, in separate runs. Tracer movement was assessed in square borosilicate glass capillaries with an inner width of 0.8 mm and wall thickness of 0.16 mm (VitroCom). Collagen type I gels (5 mg ml^{-1}) containing periosteal cells (5 million per ml) were polymerized within the capillaries, after which the capillaries were connected to the PDMS reservoirs which initiated the diffusion process resulting from a concentration gradient between the tracer saturated medium reservoir (250 μM 2-NBDG or 25 μM FL-C16 complexed to 25 μM BSA) and the tracer-free capillary. Tracer gradients within the capillaries were imaged on a confocal fluorescence laser scanning microscope (FV1000, Olympus) equipped with a UPLSAPO 10 \times air objective (NA 0.40) focused on the middle plane of the collagen gel. Focus drift was compensated using an IX81-ZDC module that focuses a 785-nm laser on the glass capillary surface to stably reproduce the focus position for each capillary position and for every acquisition time point. Images were acquired as a time series with 10-min intervals over a total period of 5 h, at 37 °C. Tracer-free collagen gels were visualized to correct the image sequences for any background fluorescence intensity. A tracer saturated collagen gel was visualized during each diffusion experiment to compensate for potential photobleaching of tracer and to normalize the gradient profiles for further processing. Image sequences were processed in ImageJ. Diffusion rates were obtained by least squares fitting an analytical solution of Fick's second diffusion law to the resulting averaged axial intensity profiles in MATLAB (MathWorks)⁵⁰.

Gene targeting

To silence *Sox9*, *Cpt1a*, *Atg5*, *Foxo1* or *Foxo3a*, we transduced cells in the presence of 8 $\mu\text{g ml}^{-1}$ polybrene (Sigma-Aldrich), with a lentivirus carrying a shRNA against SOX9[51] (Addgene plasmid repository no. 40645; multiplicity of infection (MOI) 50), CPT1a (MISSION, Sigma-Aldrich; MOI 25) or ATG5 (MISSION, Sigma-Aldrich; MOI 25), or concomitantly

with shRNAs against *Foxo1* and *Foxo3a* (MISSION, Sigma-Aldrich; each at MOI 25). To overexpress SOX9 we transduced cells, in the presence of 8 $\mu\text{g ml}^{-1}$ polybrene, with a lentivirus carrying a SOX9-overexpression plasmid[51] (Addgene plasmid repository no. 36979; MOI 150). A non-sense scrambled (Scr) shRNA sequence or empty vector was used as a negative control. After 24 h, virus-containing medium was changed to normal culture medium and 48 h later, cells were used for further experiments. Target knockdown was confirmed by western blot.

To silence expression of *Foxo* genes using CRISPR-Cas9, we transduced Cas9-expressing C3H10T1/2 cells (Cas9: Addgene plasmid repository no. 48139)⁵², with a lentivirus carrying doxycycline-inducible sgRNAs against *Foxo1* (GenBank accession number NM_019739) (5'-TTGTAAAGGTGTCTTCACGGGGG-3') and *Foxo3a* (GenBank accession number NM_019740) (5'-CATTCTGAACGCGCATGAAGCGG-3') (doxycycline-inducible plasmid: Addgene plasmid repository no. 70183)⁵³. Cells were cultured in the presence of doxycycline (250 ng ml^{-1}) for 72 h before experiments.

Quantification of active FOXO levels

Levels of active FOXO were measured using the TransAM FKHR (FOXO1) DNA-binding ELISA (Active Motif) on nuclear protein extracts, and normalized to total nuclear protein input as measured by bicinchoninic acid assay (Pierce BCA Protein Assay Kit; Thermo Scientific).

Total RNA extraction and RT-qPCR analysis

Total RNA from cultured cells was extracted using the RNeasy Mini Kit (Qiagen). Total RNA from cortical bone (femurs of eight-week-old mice, cleaned and flushed to remove bone marrow) and cartilage (growth plates dissected from the distal femur and proximal tibia of three-day-old pups) was extracted using TRIzol (Invitrogen) followed by RNA clean-up using the RNeasy Mini Kit. mRNA was reverse transcribed using Superscript II Reverse Transcriptase (Invitrogen). Reverse transcription with quantitative PCR (RT-qPCR) was performed on the 7500 Fast Real-Time PCR System (Applied Biosystems). Specific forward and reverse oligonucleotide primers were used either in conjunction with SYBR Green dye (*Cpt1a*, *Acadm*, *Acadl* and *Myod* (also known as *Myod1*)) or with FAM-TAMRA conjugated probes (all others). The following primers and probes were used: *Sox9* (GenBank accession number NM_011448): 5'-TCTGGAGGCTGCTGAACGA-3' (forward), 5'-TCCGTTCTTCACCGACTTCCT-3' (reverse), 5'-FAM-CAGCACAGAAAGACCACCC-TAMRA-3' (probe); *Col2a1* (GenBank accession number NM_031163): 5'-AGAACATCACCTACCACTGTAGAACA-3' (forward), 5'-TGACGGTCTTGCCCCACTT-3' (reverse), 5'-FAM-CCTTGCTCATCCAGGGCTCCAATG-TAMRA-3' (probe); *Acan* (GenBank accession number NM_001361500): 5'-GCATGAGAGA GGCGAATGGA-3' (forward), 5'-CTGATCTCGTAGCGATCTTTCTTCT-3' (reverse), 5'-FAM-CTGCAATTACCAGCTGCCCTTCACGT-TAMRA-3' (probe); *Runx2* (GenBank accession number NM_001146038): 5'-TACCAGCCACCGAGACCAA-3' (forward), 5'-AGAGGCTGTTTGACGCCATAG-3' (reverse), 5'-FAM-CTTGTGCCCTCTGTTGTAAATACTGCTTGCA-TAMRA-3' (probe); *Ocn* (GenBank accession number NM_007541): 5'-GGCCCTGAGTCTGACAAAGC-3' (forward), 5'-GCTCGTCACAAGCAGGGTTAA-3' (reverse), 5'-FAM-ACAGACTCCGGCGTACCTTGGAGC-TAMRA-3' (probe); *Pparg* (GenBank accession number NM_001127330): 5'-CCCAATGGTTGCTGATTACAAA-3' (forward), 5'-AATAATAAGGTGGAGATGCAGGTTCT-3' (reverse), 5'-FAM-CTGAAGCTCCAAGAATACCAAAGTGCGATC-TAMRA-3' (probe); *Myod* (GenBank accession number NM_010866): 5'-GCGCGAGTCCAGCCAGG-3' (forward), 5'-CGACTCTGGTGGTGCATCTGC-3' (reverse); *Slc2a1* (GenBank accession number NM_011400): 5'-GGGCATGTGCTTCCAGTATGT-3' (forward), 5'-ACGAGGAGCACCGTGAAGAT-3' (reverse), 5'-FAM-CAACTGTGCGGCCCTACGTCTT-TAMRA-3' (probe); *Pfkfb3* (GenBank accession number NM_001177575): Mm.PT.51.16600796 (Integrated DNA Technologies); *Ldha* (GenBank accession number NM_010699): 5'-TTCATCATTTCCCAACATTGTCAA-3' (forward),

5'-CACTGATTTCCAAGCCACGTA-3' (reverse), 5'-FAM-AGTCCACAC TGCAAGCTGCTGATCGTC-TAMRA-3' (probe); *Cpt1a* (GenBank accession number NM_013495): 5'-GCCCCATGTTGTACAGCTTCC-3' (forward), 5'-TTGGAAGTCTCCCTCCTTCA-3' (reverse); *Acadm* (GenBank accession number NM_007382): 5'-TTTGAAGACGTCAGAGTGC-3' (forward), 5'-TGGACTGTAGGTCTGGTTC-3' (reverse); *Acadl* (GenBank accession number NM_007381): 5'-TCTTTTCCTCGGAGCATGACA-3' (forward), 5'-GACCTCTACTCACTTCTCCAG-3' (reverse). Expression levels were analysed using the $2^{-\Delta\Delta C_t}$ method and were normalized for the expression of the housekeeping gene *Actb*.

RNA-seq, gene expression quantification and enrichment analysis of transcription binding motifs

In brief, total RNA was extracted from C3H10T1/2 cells seeded in six-well plates using TRIzol. Polyadenylated RNA enrichment, reverse transcription and stranded library preparation were done using the KAPA stranded mRNA-seq kit (Roche). The first 50 bases of these libraries were sequenced on a HiSeq4000 (Illumina) and mapped to the murine genome (build mm10) using TopHat v.2.1.1⁵⁴. Read counts were processed using EdgeR v.3.20.9⁵⁵ to identify genes differentially expressed between cells that were serum-starved (1% FBS) and cells that were control-treated (10% FBS). The top-100 most-significantly upregulated genes upon serum starvation (at a 1% false discovery rate, differential expression in EdgeR is assessed for each gene using an exact test analogous to Fisher's exact test, but adapted for overdispersed data⁵⁵) were analysed for motif enrichment using i-cisTarget²⁶.

Single-cell RNA-seq of mouse long bone

The single-cell RNA-seq dataset of the mouse long bone and bone marrow stroma was generated previously and detailed information on cell isolation, cell sorting, library preparation, RNA-seq and data processing is provided in the original manuscript²¹. A set of 40 genes involved in FAO and 34 genes involved in glycolysis was curated from the Gene Ontology database (<http://software.broadinstitute.org/gsea/msigdb>) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg>). Gene expression was calculated as the fraction of its unique molecular identifier (UMI; random barcode) count with respect to total UMI in the cell and then multiplied by 10,000. We denoted it as transcripts per 10,000 transcripts (TP10K).

ChIP-qPCR

ChIP-qPCR was performed as described⁵⁶. In brief, 3 h after serum deprivation, C3H10T1/2 cells were fixed using 1% formaldehyde, washed and collected by centrifugation (1,000g for 5 min at 4 °C). The pellet was resuspended in RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton-X100, 0.5% sodium deoxycholate, 1% SDS and 1% protease inhibitors), homogenized, incubated on ice for 10 min and sonicated. The samples were centrifuged (16,000g for 10 min at 4 °C) and from the supernatant sheared chromatin was used as input (1/30), and on the remainder of the chromatin immunoprecipitation was performed with an anti-FOXO1 antibody (rabbit anti-FOXO1, Abcam, ab39670) or an anti-FOXO3a antibody (rabbit anti-FOXO3a, Abcam, ab12162). After precipitation using Pierce Protein A/G Magnetic Beads (Thermo Fisher Scientific), followed by RNA and protein digestion, DNA was purified using Agencourt AMPure XP (Beckman Coulter) according to the manufacturer's instructions. RT-qPCR was performed using SYBR GreenER qPCR SuperMix Universal (Thermo Fisher Scientific) and specific primers for the *Sox9* promoter region (5'-TGTGGGCATATTG-GCTTCT-3' (forward), 5'-GGTAAACTGGGAAGACTCATGG-3' (reverse)).

Statistical analysis

All numerical results are reported as mean \pm s.e.m. Statistical significance of the difference between experimental groups was analysed by two-tailed Student's *t*-test, one-way, two-way or three-way ANOVA with Bonferroni post hoc test (as indicated in the figure legends and

source data files) using the GraphPad Prism software. Differences were considered statistically significant for $P < 0.05$. In the studies performed in cell lines in culture, all experiments were independently repeated at least three times. Experiments using primary cells were performed with at least three biological replicates. Western blots were independently repeated at least twice. Mice for experiments were randomly allocated to groups. All numerical values used for graphs and detailed statistical analysis can be found in the source data files.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The bulk RNA-seq data that support the findings of this study have been deposited in ArrayExpress with the accession number E-MTAB-7564. The single-cell RNA-seq data were generated previously²¹ and are deposited in the Gene Expression Omnibus with accession number GSE128423. A portal for exploring the entire atlas is available at https://portals.broadinstitute.org/single_cell/study/mouse-bone-marrow-stroma-in-homeostasis. Source Data for Figs. 1–4 and Extended Data Figs. 1–8 are provided with the paper. All other data supporting the findings of this study are available within the paper.

Code availability

The full code used for the computational model of bone-graft healing is available from the authors upon request. More background information on the development of the model can be found in our previous publications^{10,11}.

- Hadjantonakis, A. K., Gertsenstein, M., Ikawa, M., Okabe, M. & Nagy, A. Generating green fluorescent mice by germline transmission of green fluorescent ES cells. *Mech. Dev.* **76**, 79–90 (1998).
- Ouyang, Z. et al. *Prx1* and 3.2kb *Col1a1* promoters target distinct bone cell populations in transgenic mice. *Bone* **58**, 136–145 (2014).
- Nuyts, J. et al. Iterative reconstruction for helical CT: a simulation study. *Phys. Med. Biol.* **43**, 729–737 (1998).
- Depypere, M. et al. An iterative dual energy CT reconstruction method for a K-edge contrast material. *Proc. SPIE* **7961**, 79610M (2011).
- Vandersmissen, I. et al. Endothelial Msx1 transduces hemodynamic changes into an arteriogenic remodeling response. *J. Cell Biol.* **210**, 1239–1256 (2015).
- Maes, C. et al. Soluble VEGF isoforms are essential for establishing epiphyseal vascularization and regulating chondrocyte development and survival. *J. Clin. Invest.* **113**, 188–199 (2004).
- Stiers, P. J., van Gestel, N., Moermans, K., Stockmans, I. & Carmeliet, G. An ectopic imaging window for intravital imaging of engineered bone tissue. *J. Bone Miner. Res.* **34**, 92–102 (2018).
- Stegen, S. et al. Osteocytic oxygen sensing controls bone mass through epigenetic regulation of sclerostin. *Nat. Commun.* **9**, 2557 (2018).
- Stiers, P. J. et al. Inhibition of the oxygen sensor PHD2 enhances tissue-engineered endochondral bone formation. *J. Bone Miner. Res.* **34**, 333–348 (2019).
- McQuinn, C. et al. CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, e2005970 (2018).
- Daniëls, V. W. et al. Cancer cells differentially activate and thrive on de novo lipid synthesis pathways in a low-lipid environment. *PLoS One* **9**, e106913 (2014).
- Eyckmans, J., Lin, G. L. & Chen, C. S. Adhesive and mechanical regulation of mesenchymal stem cell differentiation in human bone marrow and periosteum-derived progenitor cells. *Biol. Open* **1**, 1058–1068 (2012).
- Ranall, M. V., Gabrielli, B. G. & Gonda, T. J. High-content imaging of neutral lipid droplets with 1,6-diphenylhexatriene. *Biotechniques* **51**, 35–42 (2011).
- Kimura, S., Noda, T. & Yoshimori, T. Dissection of the autophagosome maturation process by a novel reporter protein, tandem fluorescent-tagged LC3. *Autophagy* **3**, 452–460 (2007).
- Schoors, S. et al. Fatty acid carbon is essential for dNTP synthesis in endothelial cells. *Nature* **520**, 192–197 (2015).
- Lambrechts, D. et al. A causal relation between bioluminescence and oxygen to quantify the cell niche. *PLoS ONE* **9**, e97572 (2014).
- Guo, W. et al. Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* **148**, 1015–1028 (2012).
- Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
- Aubrey, B. J. et al. An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. *Cell Rep.* **10**, 1422–1432 (2015).
- Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).

Article

55. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
56. Stegen, S. et al. HIF-1 α promotes glutamine-mediated redox homeostasis and glycogen-dependent bioenergetics to support postimplantation bone cell survival. *Cell Metab.* **23**, 265–279 (2016).

Acknowledgements We thank K. Moermans, I. Stockmans, C. MacGillivray and H. Soled for technical assistance, A. Nagy for the CAG-eGFP mice, S. Murakami for the Col1a1-cre/ERT2 and DsRed mice, T. Yoshimori for the RFP-GFP-LC3 plasmid, R. A. Weinberg for the pLKO.1-sh-mSOX9-5 lentiviral plasmid, M. Mazzone for hypoxic glove box use, the histology core of the Harvard Department of Stem Cell and Regenerative Biology for histology services, the FACS cores of the KU Leuven and the Harvard Department of Stem Cell and Regenerative Biology for access to the flow cytometers, and the Cell Imaging Core and the Molecular Imaging and Photonics division of the KU Leuven and the Harvard Center for Biological Imaging for access to the confocal microscopes. This work was supported by grants from the Research Fund Flanders (FWO; G096414, GOA4216N and G0B3418N to G.C.), KUL grant C24/17/07 (G.C.), grants from the European Research Council (ERC 308223 to H.V.O., ERC 279100 to L.G. and ERC 269073 to P.C.), and long-term structural Methusalem funding by the Flemish Government (P.C.). N.v.G. is funded by BOF-KU Leuven GOA project 3M120209. P.J.S. is a fellow from the Agency for Innovation by Science and Technology in Flanders (IWT). S. Stegen, A.C. and Dennis L. are postdoctoral fellows of the FWO. V.W.D. is a fellow of the FWO and the Flemish League against Cancer (VLK). This work is part of Prometheus, the KU Leuven R&D Division of Skeletal Tissue Engineering.

Author contributions N.v.G. and G.C. conceived the study. N.v.G., S. Stegen, G.E., S. Schoors, P.-J.S., Dennis L., S.T. and A.S. performed the in vitro experiments. N.v.G. performed the in vivo experiments. A.C. performed the in silico experiments. V.W.D. and J.V.S. contributed to the design and execution (V.W.D.) of lipid-rescue experiments. N.B. and D.P. performed and analysed the single-cell RNA-seq experiment. M.D. and F.M. contributed to the design and execution (M.D.) of microCT analyses. R.V.L. and A.S. performed histology. P.A. contributed to the design and interpretation of autophagy experiments. N.v.G., A.C., L.G. and H.V.O. contributed to the design and interpretation of in silico experiments. Diether L. and B.T. contributed to the design, execution and interpretation of RNA-seq experiments. P.C. contributed to the design and interpretation of metabolic analyses. D.T.S. contributed to the design and interpretation of in vivo experiments. P.A., J.V.S., P.C. and D.T.S. provided reagents. N.v.G., S. Stegen and G.C. designed the experiments and interpreted data. N.v.G. and G.C. wrote the manuscript. All authors agreed on the final version of the manuscript.

Competing interests The authors declare no competing interests.

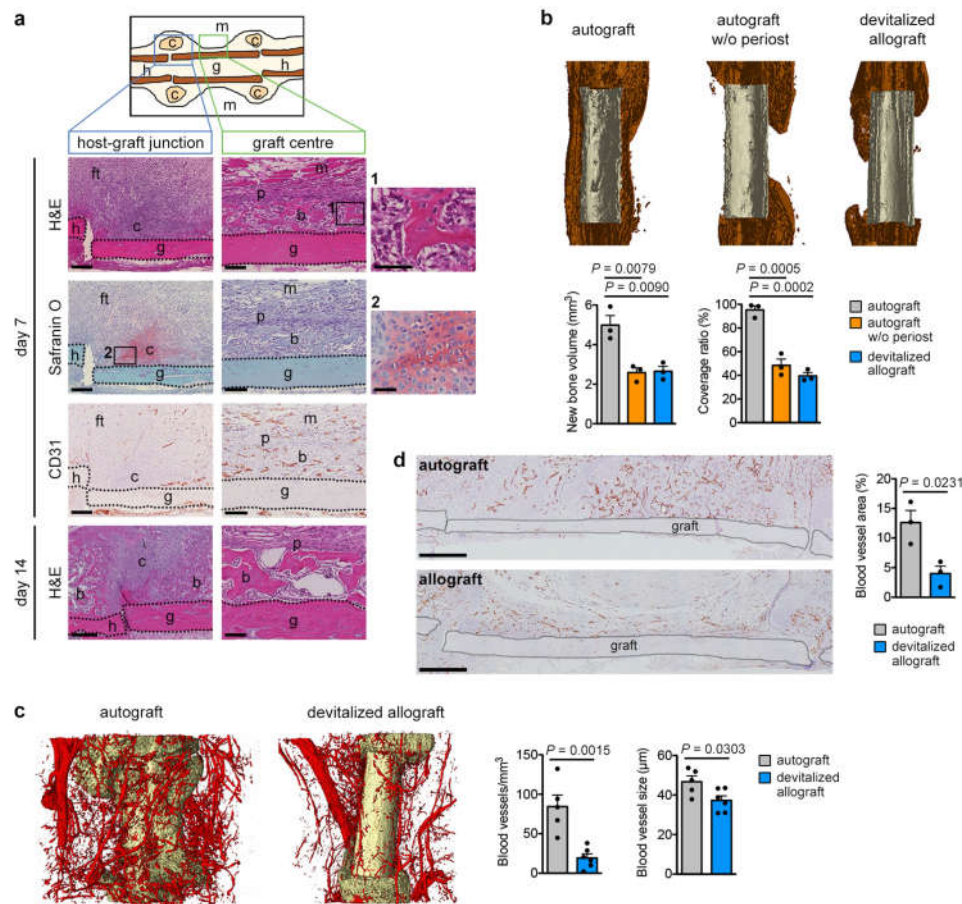
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2050-1>.

Correspondence and requests for materials should be addressed to G.C.

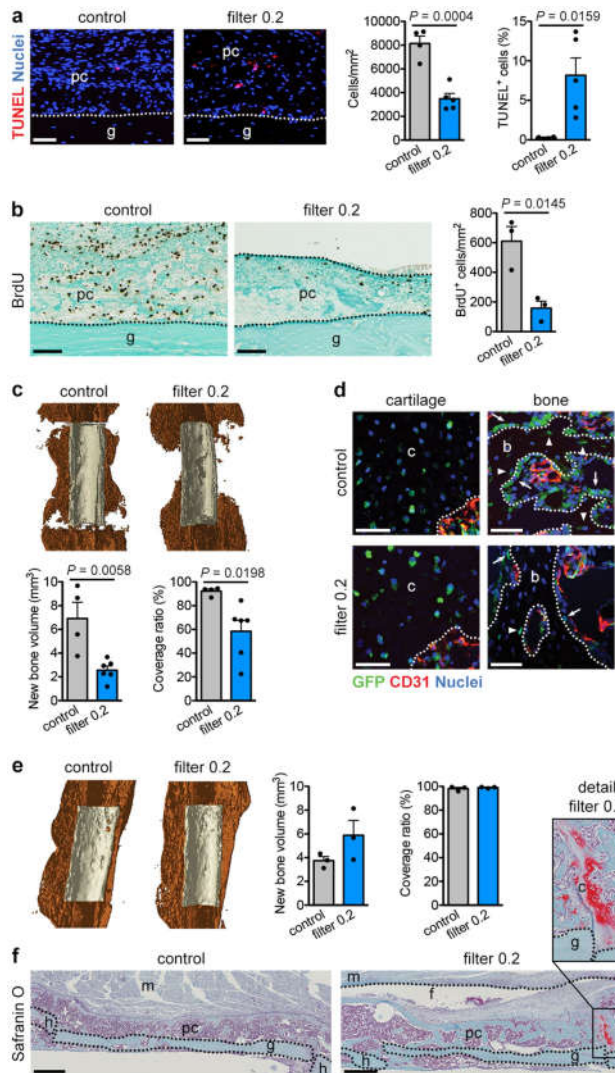
Peer review information *Nature* thanks Thomas Clemens, Michael T. Longaker and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



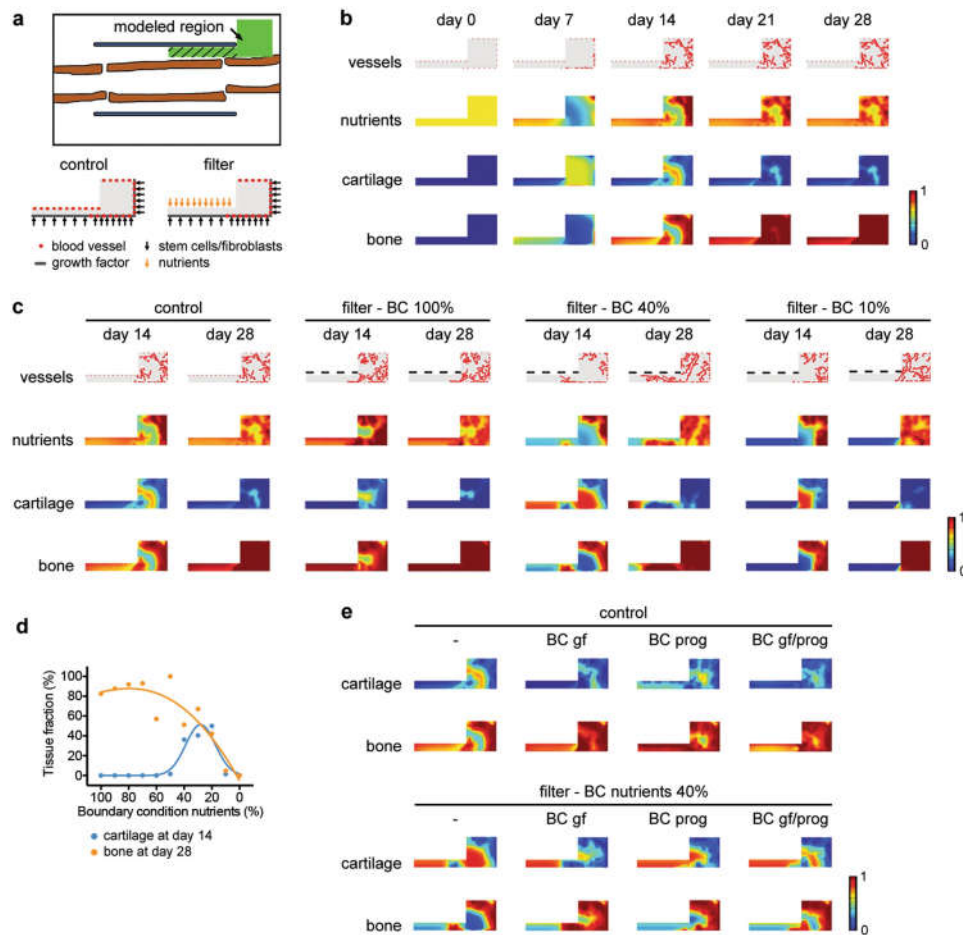
Extended Data Fig. 1 | Removal of periosteum reduces bone formation and callus vascularization. **a**, Histological characterization of the mouse bone-autograft healing model. At the host-graft junction cartilage (safranin O⁺) is formed at PFD7. Note absence of CD31⁺ blood vessels in these regions. Near the graft centre new woven bone (bright pink on H&E staining) is deposited, cartilage is absent and blood vessels are abundant. By PFD14, the cartilage at the host-graft junction is gradually being replaced by bone, while the woven bone near the graft centre appears mature (representative images of four mice). Scale bars, 200 μm in host-graft junction images, 100 μm in graft-centre images, 50 μm in magnifications. **b**, MicroCT-based visualization and quantification of newly formed bone around control autografts, autografts

from which the periosteum was removed or devitalized allografts (no living cells) at PFD28 ($n = 3$ mice). Coverage ratio represents percentage of graft surface covered by new bone. **c**, Dual-energy microCT-based visualization and quantification of vascularization in a 250-μm-wide region around autografts and allografts at PFD14 ($n = 5$ mice for autograft, $n = 6$ mice for devitalized allograft). **d**, CD31 immunohistochemical visualization and quantification of vascularization in a 250 μm-wide region around autografts and allografts at PFD14 ($n = 3$ mice). Scale bars, 500 μm. b, bone; c, cartilage; ft, fibrous tissue; g, graft; h, host; m, muscle; p, periosteum. Data are mean \pm s.e.m.; one-way ANOVA with Bonferroni post hoc test (**b**), two-tailed Student's *t*-test (**c**, **d**).



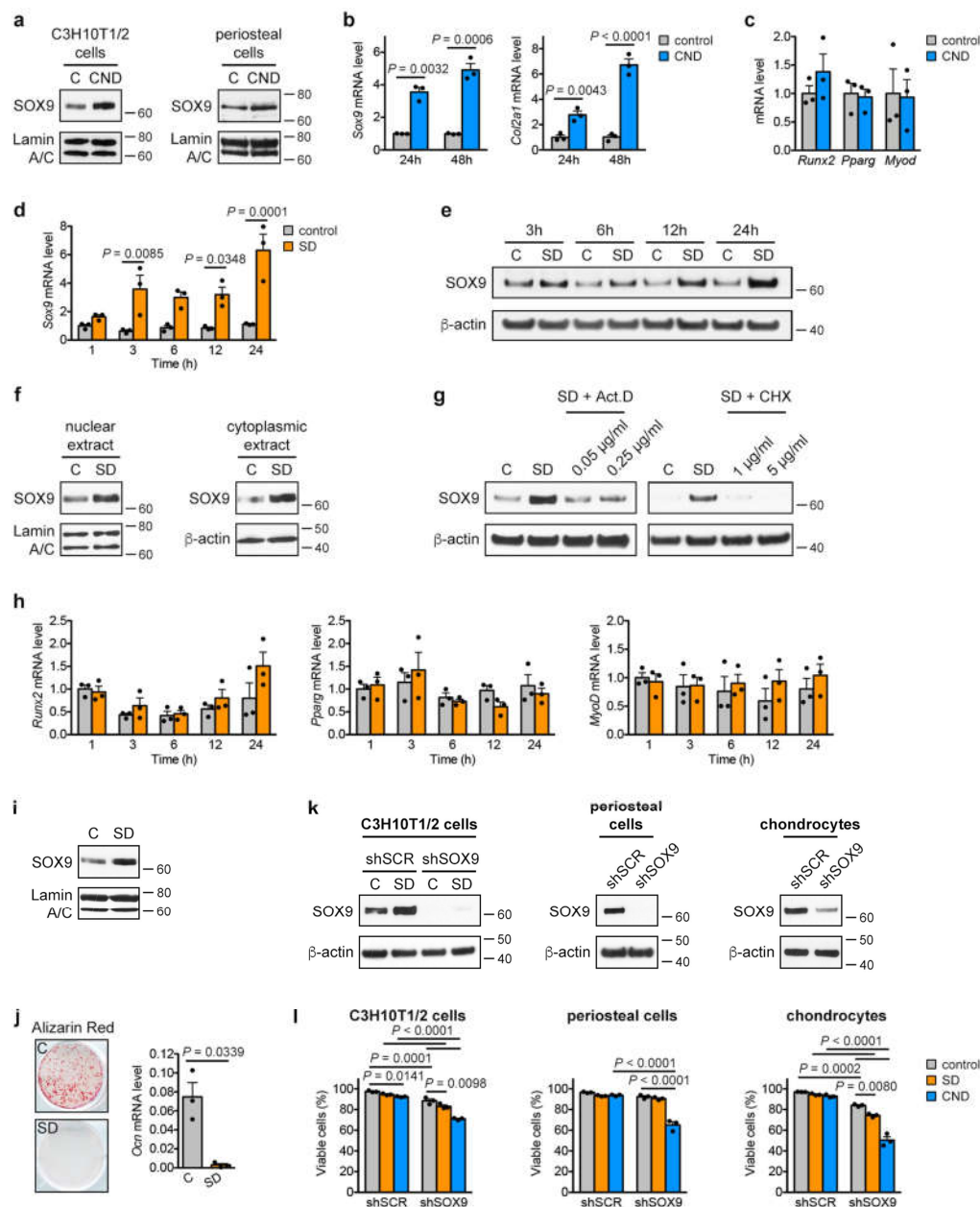
Extended Data Fig. 2 | Reducing vascularization alters but does not prevent bone healing.

a, Histological visualization and quantification of apoptotic cells (TUNEL⁺; $n = 4$ mice for control, $n = 5$ mice for filter 0.2) in the callus of grafts with or without a filter (0.2 μ m pore size) at PFD7. Scale bars, 50 μ m. **b**, Histological visualization and quantification of proliferating (BrdU⁺; $n = 3$ mice) cells in the callus of grafts with or without a filter (0.2 μ m pore size) at PFD7. Scale bars, 100 μ m. **c**, MicroCT-based visualization and quantification of newly formed bone around control grafts or grafts surrounded by a filter (0.2 μ m pore size) at PFD14 ($n = 4$ mice for control, $n = 6$ mice for filter 0.2). Coverage ratio represents percentage of graft surface covered by new bone. **d**, Cell tracing of donor periosteal cells during healing of bone grafts, derived from CAG-eGFP mice, with or without filter (0.2 μ m pore size) at PFD14 showing equal contribution of donor cells to cartilage in both conditions, but reduced contribution of donor cells to bone near the graft ends. Arrows, GFP⁺ osteoblasts; arrowheads, GFP⁺ osteocytes; representative images of three mice. Scale bars, 50 μ m. **e**, MicroCT-based visualization and quantification of newly formed bone around control grafts or grafts surrounded by a filter (0.2 μ m pore size) at PFD28 ($n = 3$ mice). **f**, Histological analysis of autografts with or without a filter (0.2 μ m pore size) at PFD28 showing comparable callus morphology and composition, although remaining cartilage islands (detail image) were seen when a filter was present but not in the callus of control grafts (representative images of three mice). Scale bars, 500 μ m. Data are mean \pm s.e.m.; two-tailed Student's *t*-test.



Extended Data Fig. 3 | In silico modelling supports a role for nutritional stress in chondrogenic commitment. Application of a previously described computational model of bone repair^{10,11} to the bone-graft healing setup. In this model, the behaviour (survival, proliferation, differentiation and tissue formation) of skeletal progenitor cells, chondrocytes, osteoblasts and fibroblasts is dependent on the local supply of nutrients by blood vessels, in addition to the presence of growth factors, extracellular matrix and the cell density. **a**, Schematic overview (top) of the modelled region shown in green. The hatched area represents the graft callus. At the start of the simulation the modelled region was filled with loose fibrous tissue matrix, growth factors, stem cells, osteoblasts, fibroblasts and nutrients, representing the fracture haematoma. Overview of the Dirichlet boundary conditions (bottom) showing the starting points of blood vessels and the sites of release of cells and growth factors (and nutrients for the condition with filter) during the healing process. **b**, Application of the model to the normal bone graft (that is, blood vessels can come from the muscle side). Heat map-based visualization of blood vessel, nutrient, cartilage and bone distribution in the modelled region at different time points shows that the model correctly predicts the spatiotemporal progression of the bone-healing process. Nutrients and tissue fractions are expressed on a non-dimensional scale ranging from 0 (absence) to 1 (saturation). **c**, **d**, Application of the model to bone-graft healing in the

presence of a filter placed in between graft and muscle (that is, blood vessels cannot come from the muscle side) with visual representation (**c**) and quantification (**d**) of the different tissue fractions in the modelled region. Quantification was performed only in the left rectangle of the modelled region, as indicated by the hatched area in **a**, representing the graft callus. The amount of nutrients that can pass through the filter (the boundary condition (BC)) was varied between 100% (the maximum amount that can be supplied by the vasculature, applied to the whole filter length, resulting in similar nutrient distributions as in the control) and 0%. When nutrient supply through the filter is set at 20–40%, the model correctly recapitulates the chondrogenic switch in the central region of the graft as observed in vivo. When nutrient supply through the filter was >40%, the cells in the central graft region differentiated directly into osteoblasts, and a supply of nutrients <20% induced massive cell death and completely prevented tissue formation and graft healing. **e**, Visual representation of the effect of additional growth factor (gf) diffusion and/or progenitor cell (prog) migration from the filter side on cartilage and bone fractions at day 14. The control situation (no filter) is shown on the top and the filter situation with a boundary condition for nutrients of 40% is shown on the bottom. No large effect of these additional boundary conditions on the healing response was observed.

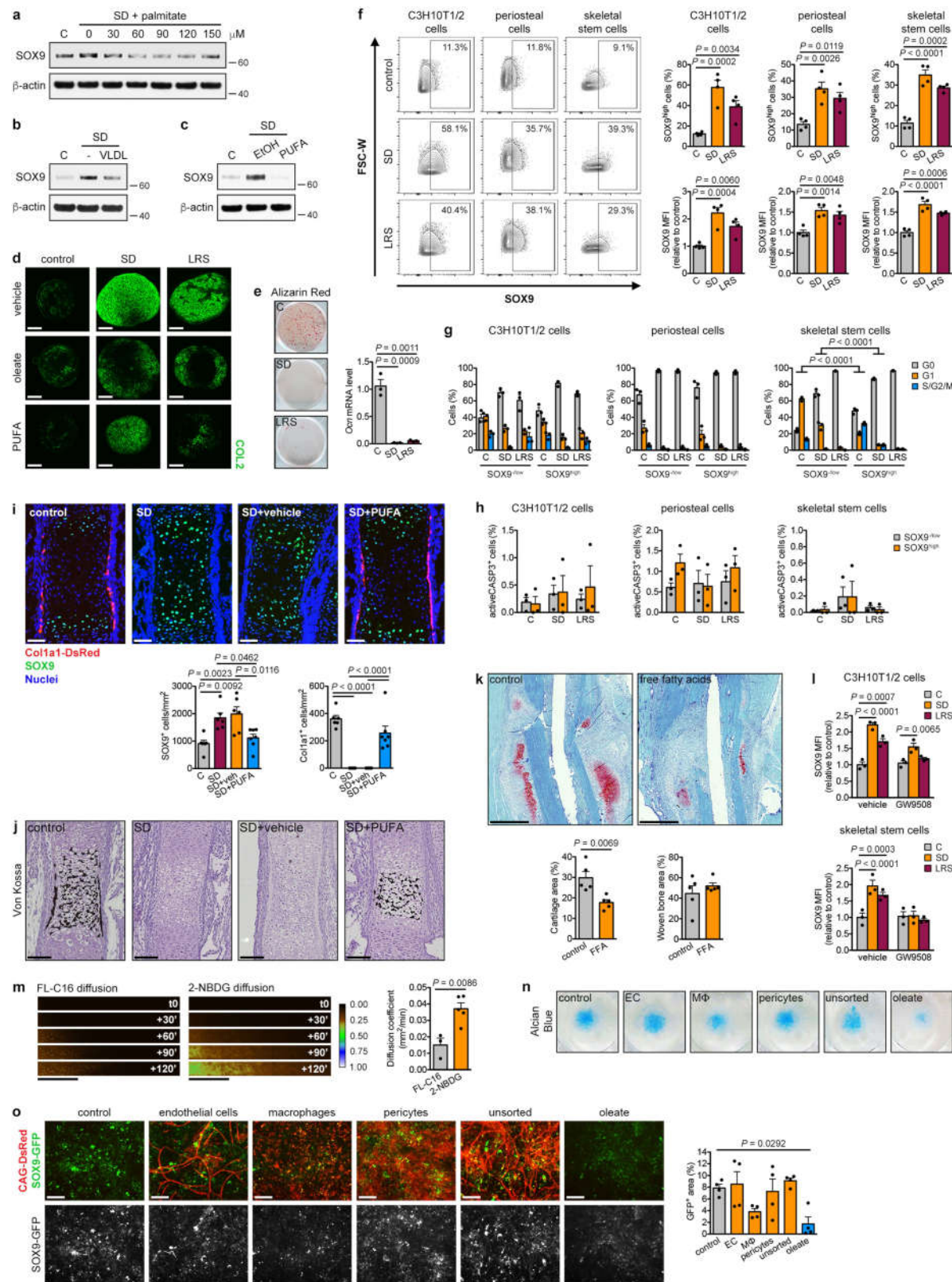


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Skeletal progenitors resist nutritional stress via induction of SOX9.

a, Immunoblot detection of nuclear SOX9 in C3H10T1/2 cells and periosteal cells exposed for 24 h to control or CND medium, with lamin A/C as loading control ($n = 2$ independent experiments). **b**, mRNA levels of *Sox9* and *Col2a1* in periosteal cells exposed for the indicated times to control or CND medium (relative to control; $n = 3$ biologically independent samples). **c**, mRNA levels of runt-related transcription factor 2 (*Runx2*; osteogenic lineage), peroxisome proliferator-activated receptor γ (*Pparg*; adipogenic lineage) and *Myod* (myogenic lineage) in periosteal cells exposed for 48 h to control or CND medium (relative to control; $n = 3$ biologically independent samples). **d**, mRNA levels of *Sox9* in C3H10T1/2 cells exposed for the indicated times to control or SD medium (relative to control, $n = 3$ independent experiments). **e**, Immunoblot detection of total SOX9 in C3H10T1/2 cells exposed for different durations to control or SD medium, with β -actin as loading control ($n = 2$ independent experiments). **f**, Immunoblot detection of nuclear and cytoplasmic SOX9 in C3H10T1/2 cells exposed for 6 h to control or SD medium, with lamin A/C or β -actin as loading control ($n = 2$ independent experiments). **g**, Immunoblot detection of SOX9 in total cell protein extracts of C3H10T1/2 cells exposed for 6 h to control medium, SD medium or SD medium supplemented with different concentrations of the transcription inhibitor actinomycin D (Act. D) or the translation inhibitor cycloheximide (CHX). Detection of β -actin was used as loading control ($n = 2$ independent

experiments). **h**, mRNA levels of *Runx2*, *Pparg* and *Myod* in C3H10T1/2 cells exposed for the indicated times to control or SD medium (relative to control, $n = 3$ independent experiments). **i**, Immunoblot detection of nuclear SOX9 in periosteal cells exposed for 24 h to control or SD medium with lamin A/C as loading control ($n = 3$ biologically independent samples). **j**, Osteogenic differentiation of periosteal cells in control or SD medium, assessed by visualization of mineral deposits (alizarin red staining) and quantification of *Ocn* mRNA levels (relative to *Actb*, $n = 3$ biologically independent samples). **k**, Immunoblot detection of SOX9 in total cell protein extracts of C3H10T1/2 cells (in control or SD medium), periosteal cells and growth plate-derived chondrocytes transduced with sh*Sox9* or shScr, with β -actin as loading control. A longer exposure time was used for SOX9 detection in C3H10T1/2 cells and periosteal cells compared with chondrocytes in order to visualize any remaining protein in the shSOX9 conditions ($n = 2$ independent experiments for C3H10T1/2 cells, $n = 3$ biologically independent samples for periosteal cells, growth plate-derived chondrocytes). **l**, Quantification of cell viability of C3H10T1/2 cells, periosteal cells and growth plate-derived chondrocytes transduced with sh*Sox9* or shSCR, after 72 h of exposure to control, SD or CND medium ($n = 3$ independent experiments for C3H10T1/2 cells, $n = 3$ biologically independent samples for periosteal cells, growth plate-derived chondrocytes). Data are mean \pm s.e.m.; two-way ANOVA with Bonferroni post hoc test (**b**, **d**, **h**, **l**), two-tailed Student's *t*-test (**c**, **j**). For gel source data, see Supplementary Fig. 1.

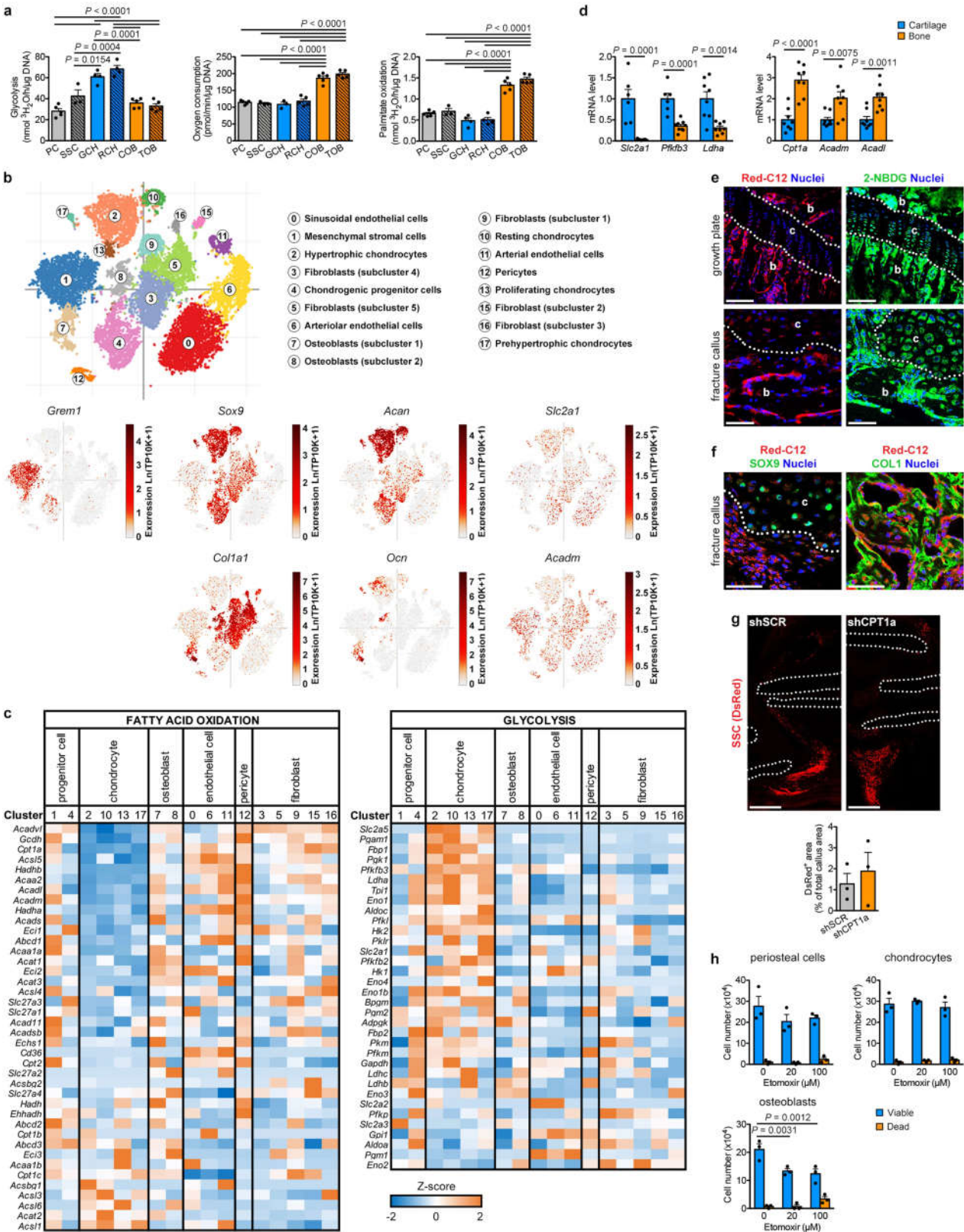


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Reduced lipid availability favours chondrogenesis over osteogenesis.

a–c, Immunoblot detection of total SOX9 in C3H10T1/2 cells exposed for 6 h to control medium, SD medium or SD medium supplemented with increasing concentrations of palmitate (**a**), VLDL (**b**) or PUFA (**c**). Detection of β -actin was used as loading control. EtOH was used as a vehicle control in **a** and **c** ($n=2$ independent experiments). **d**, Histological visualization (by immunofluorescence for COL2) of chondrogenic differentiation of periosteal cells in pellet cultures in control, SD or LRS medium supplemented with vehicle (EtOH), oleate or PUFA (representative images of $n=2$ independent experiments). Scale bars, 100 μm . **e**, Osteogenic differentiation of periosteal cells in control, SD or LRS medium, assessed by visualization of mineral deposits (alizarin red staining) and quantification of *Ocn* mRNA levels (relative to *Actb*, $n=3$ biologically independent samples). **f**, Flow cytometric detection and quantification of the percentage of SOX9^{high} cells and total SOX9 levels in C3H10T1/2 cells, periosteal cells and skeletal stem cells exposed for 24 h to control, SD or LRS medium ($n=4$ independent experiments for C3H10T1/2 cells, $n=4$ biologically independent samples for periosteal cells, skeletal stem cells). Gating for SOX9^{high} cells was set to have approximately 10% SOX9^{high} cells in control conditions in each cell type. **g, h**, Flow cytometric quantification of cell cycle (**g**) and apoptosis (**h**) in SOX9^{low} and SOX9^{high} subpopulations of C3H10T1/2 cells, periosteal cells and skeletal stem cells exposed for 24 h to control, SD or LRS medium ($n=3$ independent experiments for C3H10T1/2 cells, $n=3$ biologically independent samples for periosteal cells, skeletal stem cells). **i**, Histological visualization and quantification of early chondrogenic (SOX9⁺) and osteogenic (Col1a1-DsRed⁺) cells in metatarsals cultured for one week in control medium, SD medium or SD medium supplemented with PUFA or vehicle (EtOH) ($n=6$

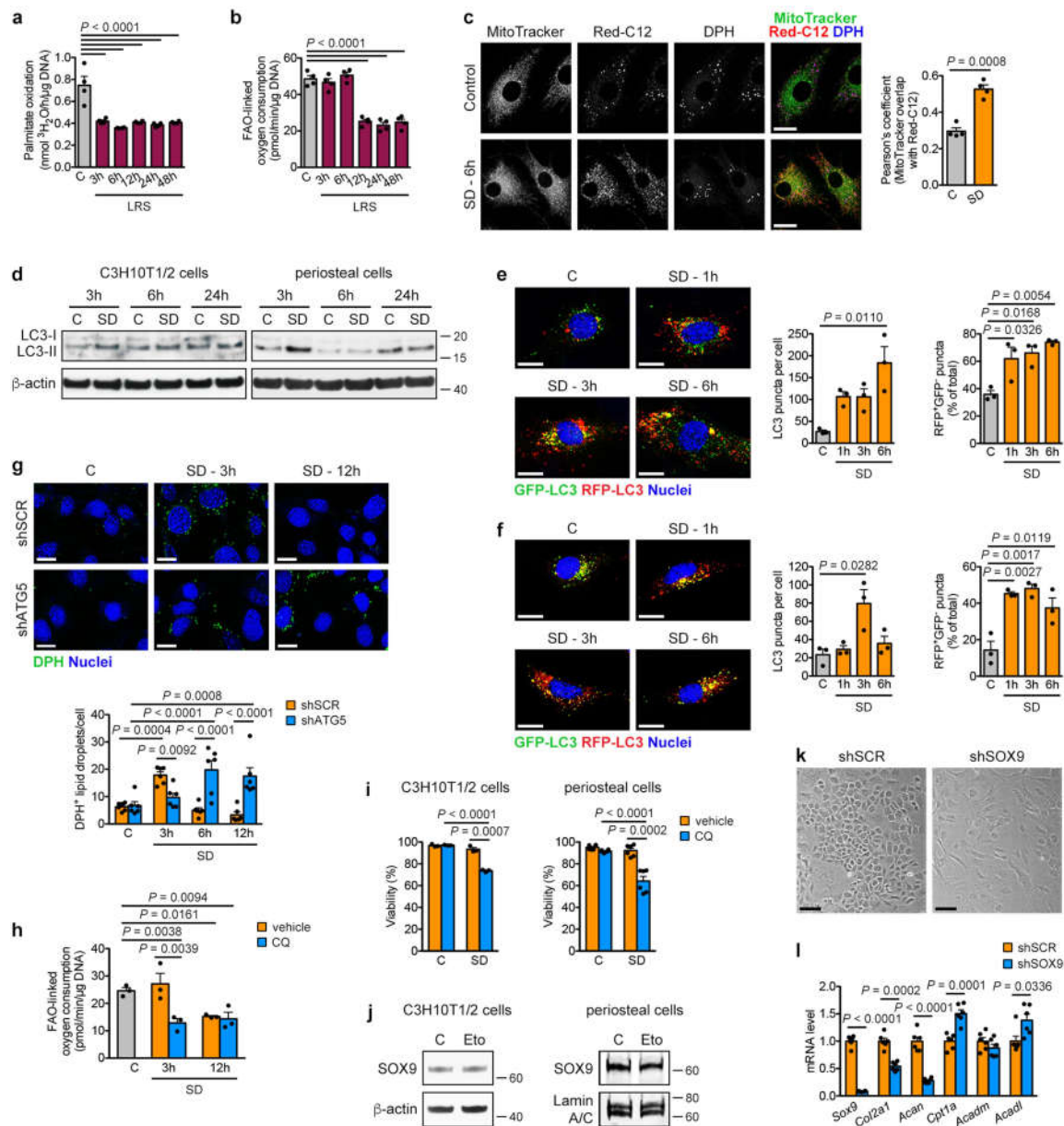
biologically independent samples for control, SD and SD + vehicle, $n=7$ biologically independent samples for SD + PUFA). Scale bars, 50 μm . **j**, Histological visualization of mineralization by Von Kossa staining in metatarsals cultured for one week in control medium, SD medium or SD medium supplemented with vehicle or PUFA (representative images of $n=6$ biologically independent samples for control, SD and SD + vehicle, $n=7$ biologically independent samples for SD + PUFA). Scale bars, 100 μm . **k**, Histological visualization (safranin O staining) and quantification of cartilage and woven bone in the callus at PFD7 of mice treated daily with free fatty acids (FFA; 20 μl corn oil) or sham injection (saline) at the fracture site ($n=5$ mice). Scale bars, 500 μm . **l**, Flow cytometric quantification of total SOX9 levels in C3H10T1/2 cells or skeletal stem cells exposed for 24 h to control, SD or LRS medium supplemented with 100 μM GW9508 or vehicle (DMSO) ($n=3$ independent experiments for C3H10T1/2 cells, $n=3$ biologically independent samples for skeletal stem cells). **m**, Visualization and quantification of diffusion of a fluorescent fatty acid (FL-C16) and fluorescent glucose (2-NBDG) in collagen gels seeded with periosteal cells (5×10^6 per ml) ($n=3$ biologically independent samples for FL-C16, $n=5$ biologically independent samples for 2-NBDG). Scale bars, 500 μm . **n, o**, Visualization of alcian blue staining (**n**) and visualization and quantification of Sox9 expression (**o**) in micromass co-cultures of periosteal cells from Sox9-GFP mice and sorted cell populations from skeletal muscle of CAG-DsRed mice, after nine days in chondrogenic SD medium ($n=4$ biologically independent samples). Addition of oleate was used as positive control. Scale bars, 100 μm . EC, endothelial cell, M Φ , macrophage. Data are mean \pm s.e.m.; one-way ANOVA (**e, f, i, o**), two-way ANOVA (**h, l**) or three-way ANOVA (**g**) with Bonferroni post hoc test, two-tailed Student's *t*-test (**k, m**). For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Chondrocytes do not depend on FAO. **a**, Quantification of glycolytic rate, oxygen consumption and palmitate oxidation in periosteal cells (PC, $n = 5$ biologically independent samples), skeletal stem cells (SSC, $n = 3$ biologically independent samples), growth plate-derived chondrocytes (GCH, $n = 3$ biologically independent samples for oxygen consumption, $n = 4$ biologically independent samples for glycolysis and palmitate oxidation), rib chondrocytes (RCH, $n = 5$ biologically independent samples for oxygen consumption, $n = 4$ biologically independent samples for glycolysis and palmitate oxidation), calvarial osteoblasts (COB, $n = 5$ biologically independent samples) and trabecular osteoblasts (TOB, $n = 5$ biologically independent samples). **b**, t -Distributed stochastic neighbour embedding (t -SNE) plot of 20,896 non-haematopoietic cells (mixed bone and bone marrow fractions, $n = 6$ mice) based on single-cell RNA-seq data, annotated post hoc and coloured by clustering (top) or by expression (ln(TP10K)) of selected genes (bottom). **c**, Expression (row-wide Z score of ln of average TP10K; single-cell RNA-seq) of FAO- and glycolysis-related genes (rows) in the cells of each cluster (columns). **d**, RT-qPCR analysis of genes involved in glycolysis (*Glut1* (also known as *Slc2a1*), *Pfkfb3* and *Ldha*; $n = 6$ independent samples for *Glut1* and *Pfkfb3* in cartilage, $n = 9$ independent samples for *Glut1* and *Pfkfb3* in bone, $n = 8$

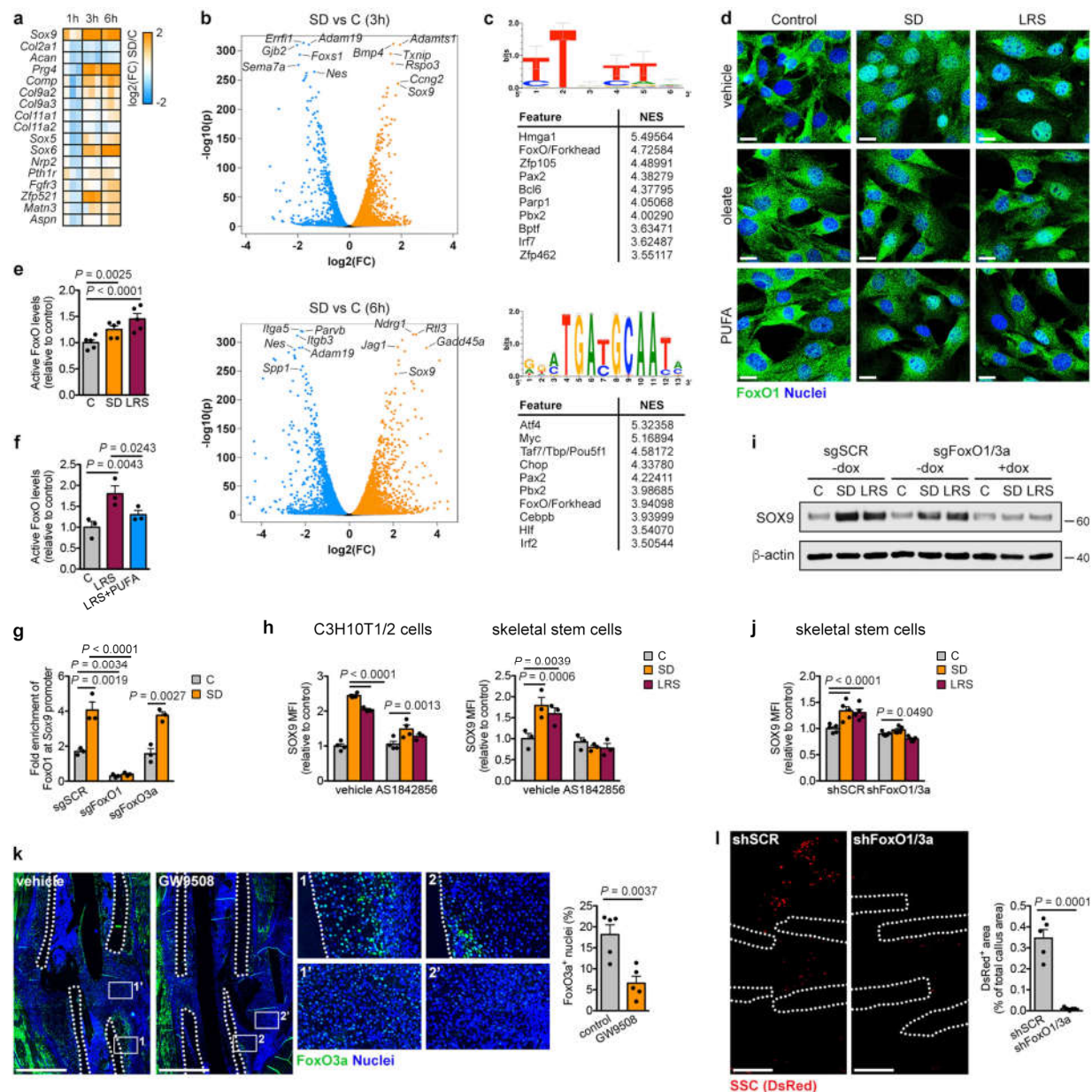
independent samples for *Ldha*) and FAO (*Cpt1a*, *Acadm* and *Acadl*; $n = 8$ independent samples) in mouse growth plate cartilage and cortical bone biopsies (relative to *Actb*). **e**, Analysis of adjacent histological sections of a growth plate and fracture callus (PFD7) of mice injected intravenously with a fluorescent fatty acid (Red-C12) or glucose (2-NBDG) (representative images of $n = 3$ mice). Scale bars, 100 μ m in growth plate images, 50 μ m in fracture callus images. **f**, Immunofluorescence analysis of a fracture callus (PFD7) of a mouse injected intravenously with a fluorescent fatty acid (Red-C12) and stained for SOX9 (left; cartilage area shown) or COL1 (right; trabecular bone area shown) (representative images of $n = 3$ mice). Scale bars, 50 μ m. **g**, Histological visualization and quantification at PFD7 of CAG-DsRed⁺ skeletal stem cells (SSC), transduced with sh*Cpt1a* or shScr and transplanted at the fracture site on PFD0 ($n = 3$ mice). Dotted lines delineate cortical bone ends. **h**, Quantification of number of live and dead cells in cultures of periosteal cells, growth plate-derived chondrocytes and calvarial osteoblasts after 48 h of exposure to etomoxir ($n = 3$ biologically independent samples). Data are mean \pm s.e.m.; one-way (**a**) or two-way (**h**) ANOVA with Bonferroni post hoc test, two-tailed Student's t -test (**d**, **g**).



Extended Data Fig. 7 | Changes in FAO and autophagy after lipid deprivation.

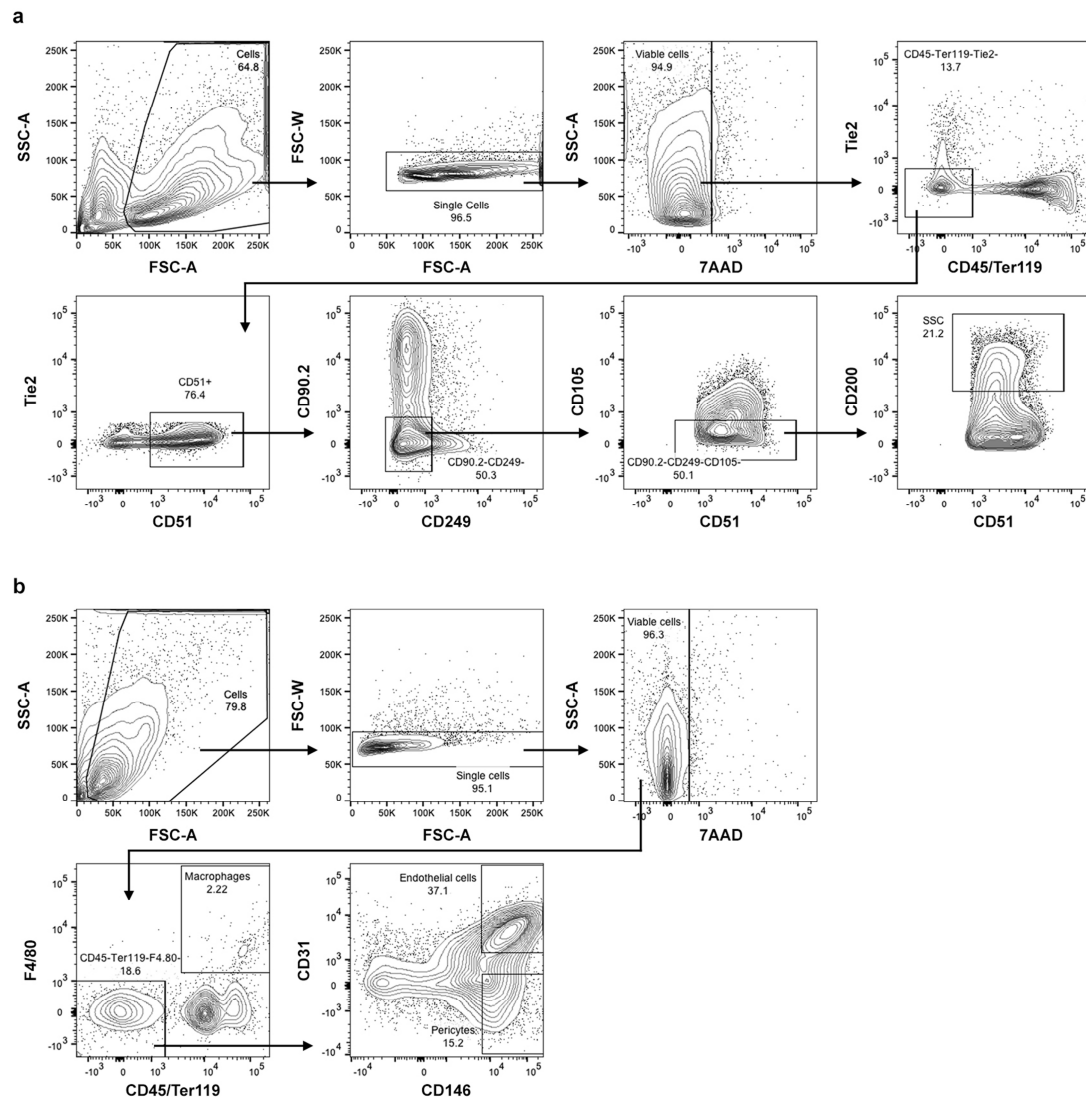
a, Oxidation of extracellularly added palmitate by periosteal cells in control medium or at different times in LRS medium ($n = 4$ biologically independent samples). **b**, Quantification of FAO-linked OCR in periosteal cells in control medium or at different times in LRS medium ($n = 4$ biologically independent samples). **c**, Confocal microscopy of periosteal cells labelled with Red-C12 (fluorescent fatty acid, red) and stained with MitoTracker (mitochondria, green) and DPH (lipid droplets, blue) shows increased colocalization (as quantified by Pearson's correlation coefficient) of MitoTracker and Red-C12 after exposure of cells for 6 h to SD medium ($n = 4$ biologically independent samples). Scale bars, 20 μm . **d**, Immunoblot detection of LC3 in total cell protein extracts of C3H10T1/2 cells and periosteal cells exposed for different times to control or SD medium, with β -actin as loading control. Note increased conversion of LC3-I to LC3-II at early time points, indicative of activation of autophagy ($n = 2$ independent experiments). **e**, **f**, Confocal microscopy of C3H10T1/2 cells (**e**; $n = 3$ independent experiments) or periosteal cells (**f**; $n = 3$ biologically independent samples), expressing an RFP-GFP-LC3 tandem construct, shows activation of autophagy with time upon serum deprivation, evidenced by increased total number of LC3 puncta per cell and higher percentage of RFP⁺GFP⁺ puncta. Scale bars, 20 μm . **g**, Confocal microscopy-based visualization (top) and quantification (bottom) of C3H10T1/2 cells, stained with the neutral lipid dye DPH to reveal lipid-droplet dynamics at different time points after SD. Cells were transduced with shAtg5 to inhibit

autophagy or shScr as a control ($n = 6$ independent experiments). Scale bars, 20 μm . **h**, Quantification of FAO-linked OCR in periosteal cells in control medium or at different times after serum deprivation, treated with 10 μM chloroquine (CQ) or vehicle ($n = 3$ biologically independent samples). **i**, Quantification of cell viability of C3H10T1/2 cells and periosteal cells after 72 h of exposure to control or SD medium in the presence or absence of 50 μM (C3H10T1/2 cells) or 10 μM (periosteal cells) CQ ($n = 3$ independent experiments for C3H10T1/2 cells, $n = 3$ biologically independent samples for periosteal cells). **j**, Immunoblot detection of total SOX9 in C3H10T1/2 cells and nuclear SOX9 in periosteal cells exposed for 6 h (C3H10T1/2 cells) or 24 h (periosteal cells) to control medium (with DMSO as vehicle control) or medium supplemented with 100 μM etomoxir (Eto), with β -actin or lamin A/C as loading control ($n = 2$ independent experiments for C3H10T1/2 cells, $n = 3$ biologically independent samples for periosteal cells). **k**, Cell morphology of growth plate-derived chondrocytes transduced with shSox9 or shScr (representative images of six biologically independent samples). Scale bar, 100 μm . **l**, RT-qPCR analysis of genes involved in chondrogenesis (*Sox9*, *Col2a1* and *Acan*) and FAO (*Cpt1a*, *Acadl* and *Acat*) in growth plate-derived chondrocytes transduced with shSox9 or shScr (relative to shScr, $n = 6$ biologically independent samples). Data are mean \pm s.e.m.; one-way ANOVA (**a**, **b**, **e**, **f**) or two-way ANOVA (**g**–**i**) with Bonferroni post hoc test, two-tailed Student's *t*-test (**c**, **l**). For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 8 | Lipids regulate SOX9 through FOXO signalling. **a**, Heat map showing differential expression of cartilage-related genes in C3H10T1/2 cells exposed for different times to SD versus control medium, as determined by RNA-seq ($n=3$ replicates). **b**, Volcano plot showing significantly enriched and depleted mRNAs in C3H10T1/2 cells exposed for 3 or 6 h to SD versus control medium, as determined by RNA-seq ($n=3$ replicates). **c**, Top 10 most significantly enriched transcription factor motifs with normalized enrichment scores (NES) in C3H10T1/2 cells exposed for 3 h (left) or 6 h (right) to SD versus control medium, as determined by i-cisTarget analysis on the 100 most significantly increased mRNAs ($n=3$ replicates). Motif shown on top is the *Hmga1* motif for 3 h and the *Atf4* motif for 6 h. **d**, Confocal microscopy of C3H10T1/2 cells stained for FOXO1 after exposure of cells for 3 h to SD or LRS medium in the presence of vehicle (EtOH), oleate (60 μ M) or PUFA (representative images of two independent experiments). Scale bars, 20 μ m. **e**, Nuclear FOXO activity in C3H10T1/2 cells exposed for 3 h to control, SD or LRS medium ($n=5$ independent experiments). **f**, Nuclear FOXO activity in skeletal stem cells exposed for 3 h to control medium, LRS medium or LRS medium supplemented with PUFA ($n=3$ biologically independent samples). EtOH was used as vehicle control. **g**, Occupancy of FOXO1 at the *Sox9* promoter of Cas9-expressing C3H10T1/2 cells transduced with *sgFoxo1*, *sgFoxo3a* or *sgScr*, exposed for 3 h to control or SD medium, as determined by ChIP-qPCR ($n=3$ independent experiments). **h**, Flow cytometric quantification of total

SOX9 levels in C3H10T1/2 cells ($n=4$ independent experiments for control and serum deprivation, $n=3$ independent experiments for LRS) and skeletal stem cells ($n=3$ biologically independent samples) exposed for 24 h to control, SD or LRS medium supplemented with 1 μ M AS1842856 or vehicle (DMSO). **i**, Immunoblot detection of total SOX9 in Cas9-expressing C3H10T1/2 cells transduced with inducible *sgFoxo1* and *sgFoxo3a* (*sgFoxo1/3a*) or with *sgScr*, exposed for 6 h to control, SD or LRS medium in the presence or absence of doxycycline (dox; 250 μ g ml^{-1}), with β -actin as loading control ($n=2$ independent experiments). **j**, Flow cytometric quantification of total SOX9 levels in skeletal stem cells transduced with *shFoxo1* and *shFoxo3a* (*shFoxo1/3a*) or with *shScr*, exposed for 24 h to control, SD or LRS medium ($n=5$ biologically independent samples). **k**, Histological visualization and quantification of FOXO3a-expressing cells in the fracture callus at PFD7 of mice treated daily with GW9508 (10 nmol) or vehicle (0.2% DMSO in saline) at the fracture site ($n=5$ mice). Scale bars, 500 μ m. Dotted lines delineate cortical bone ends. **l**, Histological visualization and quantification in the fracture callus at PFD7 of CAG-DsRed⁺ skeletal stem cells (SSC), transduced with *shFoxo1/3a* or *shScr* and transplanted at the fracture site on PFD0 ($n=5$ mice). Dotted lines delineate cortical bone ends. Data are mean \pm s.e.m.; one-way ANOVA (**e**, **f**), two-way ANOVA (**g**, **h**, **j**) with Bonferroni post hoc test, two-tailed Student's *t*-test (**k**, **l**). For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 9 | Flow cytometry gating for cell sorting. a, Contour plots showing the gating strategy for the identification and isolation of skeletal stem cells from long bones of newborn mice. **b,** Contour plots showing the

gating strategy for the identification and isolation of macrophages, endothelial cells and pericytes from skeletal muscle of adult mice.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

qRT-PCR: StepOne Real-Time PCR software 2.3
Flow cytometry: BD FACSDiva 8.0
Seahorse flux analyzer: XF Reader 1.8.1.1
scintillation counting: QuantaSmart TM 4.0 Perkin Elmer
Imaging: Zeiss Zen 2.5, Zeiss AxioVision 4.9.1
Computer Modeling: MatLab R2012a/R2013a

Data analysis

CT: CT Analyzer 1.16.4.1, CT Vol 2.3.2.0, MeVisLab 2.6.2
Flow cytometry: FlowJo 10.5.3
Statistics: GraphPad Prism 8.1.2
Graphs: GraphPad Prism 5.0
Image analysis: ImageJ/FIJI 2.0.0, CellProfiler 3.1.8
Image preparation: Photoshop CS5
RNAseq/scRNAseq analysis: TopHat 2.1.1, R statistical software 3.5.1, EdgeR 3.20.9, i-cisTarget

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The bulk mRNA sequencing data that support the findings of this study have been deposited in ArrayExpress with the accession number E-MTAB-7564 (<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7564>). The single cell RNA sequencing data were generated previously and are deposited in GEO (GSE128423, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128423>). A portal for exploring the entire atlas is available (https://portals.broadinstitute.org/single_cell/study/mouse-bone-marrow-stroma-in-homeostasis). All other data supporting the findings of this study are available within the paper.

Figures 2,4 and Extended Data Figures 4,5,7,8 have associated raw data, provided as Supplemental Information Figure 1 for uncropped blot pictures

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical tests were used to pre-determine sample size, but sample size was chosen based on previous experiments and comparable studies in literature. Sample size for each experiment is indicated in the legend.
Data exclusions	No data was excluded.
Replication	In the studies performed in cell lines in culture, all experiments were independently repeated at least three times. Experiments using primary cells were performed with at least three biological replicates. Western Blots were independently repeated at least twice. All attempts at replication were successful
Randomization	Mice for experiments were randomly allocated to groups.
Blinding	Blinding was widely used in the study. Data collection and analysis, such as immunostaining, qRT-PCR, and Western blot were frequently performed by participants other than the experiment designer. During these data collection and analysis steps, all participants were routinely blinded to group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Antibodies for flow cytometry:
 PE/Cy7 anti-mouse CD45: BioLegend, #103114, clone 30-F11, lot B243728, 1/200 dilution
 PE/Cy7 anti-mouse TER-119: BioLegend, #116222, clone TER-119, lot B251241, 1/200 dilution
 APC anti-mouse CD202b (Tie-2, CD202): BioLegend, #124010, clone TEK4, lot B231548, 1/200 dilution
 PE anti-mouse CD202b (Tie-2, CD202): BioLegend, #124008, clone TEK4, lot B207408, 1/200 dilution

APC anti-mouse CD105: BioLegend, #120414, clone MJ7/18, lot B204640, 1/200 dilution
 Pacific Blue anti-mouse CD105: BioLegend, #120412, clone MJ7/18, lot B245562, 1/200 dilution
 APC anti-mouse CD90.2: BioLegend, #105312, clone 30-H12, lot B208842, 1/200 dilution
 FITC anti-mouse Ly-51 (6C3): BioLegend, #108305, clone 6C3, lot B218198, 1/200 dilution
 PerCP-eFluor 710 anti-mouse CD200: eBioscience, #46-5200-82, clone OX90, lot 4298111, 1/200 dilution
 Biotin anti-mouse CD51: BD Pharmingen, #551380, clone RMV-7, lot 3301919, 1/200 dilution
 PE anti-Ki67: BD Pharmingen, #556027, clone B56, lot 5113743, 1/10 dilution
 PE anti-active Caspase 3: BD Pharmingen, #550821, clone C92-605, lot 25660, 1/100 dilution
 AlexaFluor 647 anti-SOX9: Cell Signaling Technology, #71273, clone D8G8H, lot 1, 1/100 dilution
 APC anti-mouse CD146: BioLegend, #134712, clone ME-9F1, lot B268897, 1/200 dilution
 Pacific Blue anti-mouse F4/80: BioLegend, #123124, clone BM8, lot B217178, 1/200 dilution
 PerCP/Cy5.5 anti-mouse CD31: BioLegend, #102522, clone MEC13.3, lot 4052815, 1/200 dilution

Primary antibodies for immunohistochemical analysis:

BrdU: Bio-Rad, #OBT0030, clone BU 1/75-ICR1, Rat, lot 0512, 1/500 dilution
 CD31: BD Biosciences, 550274, clone MEC 13.3, Rat, lot 7292994, 1/50 dilution
 type I collagen: Novus Biologicals, clone NB600-408, polyclonal, rabbit, lot 40267, 1/100 dilution
 type II collagen: Merck, MAB8887, monoclonal, mouse, clone 6B3, lot 2933390, 1/200 dilution
 SOX9: Novus Biologicals, NBP1-85551, polyclonal, rabbit, lot B113838, 1/200 dilution
 CPT1a: Cell Signaling Technology, #12252, clone D3B3, rabbit, lot 1, 1/50 dilution
 GLUT1: Cell Signaling Technology, #12939, clone D3J3A, rabbit, lot 1, 1/100 dilution
 FoxO3a: Cell Signaling Technology, #2497, clone 75D8, rabbit, lot 5, 1/100 dilution

Secondary antibodies for immunohistochemical analysis:

Biotin anti-mouse: Dako, #E0433, goat, lot 00062137
 Fluorescein anti-mouse: Sigma-Aldrich, #F0257, goat, lot SLBV6490
 Biotin anti-rat: BD Biosciences, #559286, goat, lot 6321784
 Biotin anti-rat: Dako, #E0468, rabbit, lot 00043220
 Biotin anti-rabbit: Dako, #E0432, goat, lot 20027287
 Cy3 anti-rabbit: Jackson ImmunoResearch, #111-165-003, goat, lot 84241
 AlexaFluor 546 anti-rabbit: Invitrogen, #A-11010, goat, lot 1904467
 AlexaFluor 488 anti-rabbit: Invitrogen; #A-11034, goat, lot 1937195

Primary antibodies for immunocytochemical analysis:

FoxO1: Cell Signaling Technology, #2880, clone C29H4, rabbit, lot 11, 1/100 dilution
 FoxO3a: Cell Signaling Technology, #2497, clone 75D8, rabbit, lot 5, 1/100 dilution

Secondary antibodies for immunocytochemical analysis:

AlexaFluor 488 anti-rabbit: Invitrogen; #A-11034, goat, lot 1937195

Primary antibodies for Western Blot:

SOX9: Novus Biologicals, NBP1-85551, polyclonal, rabbit, lot B113838, 1/2000 dilution
 FoxO1: Cell Signaling Technology, #2880, clone C29H4, rabbit, lot 11, 1/1000 dilution
 FoxO3a: Cell Signaling Technology, #2497, clone 75D8, rabbit, lot 5, 1/1000 dilution
 LC3B: Cell Signaling Technology, #3868, clone D11, rabbit, lot 11, 1/500 dilution
 β -actin: Sigma-Aldrich, #A5441, clone AC-15, mouse, lot 026M, 1/10000 dilution
 Lamin A/C: Santa Cruz Biotechnology, #sc-376248, clone E-1, mouse, lot C1412, 1/5000 dilution

Secondary antibodies for Western blot analysis:

HRP anti-mouse: Dako, #P0161, rabbit, lot 00095192
 HRP anti-rabbit: Dako, #P0448, goat, lot 00094764

Primary antibodies for ChIP-qPCR:

FoxO1: Abcam, #ab39670, rabbit, lot GR3192176-1, 1/250 dilution
 FoxO3a: Abcam, #ab12162, rabbit, lot GR226465-14, 1/250 dilution

Validation

All antibodies were obtained from indicated commercial vendors with ensured quality. In addition, all the antibodies have been used in multiple experiments to detect intended proteins in control samples with expected molecular weight to validate their effectiveness in our study

Antibodies for flow cytometry:

PE/Cy7 anti-mouse CD45: BioLegend, #103114, clone 30-F11
 RRID: AB_312979

Validated by the manufacturer by flow cytometry on C57BL/6 mouse splenocytes, 46 citations

PE/Cy7 anti-mouse TER-119: BioLegend, #116222, clone TER-119
 RRID: AB_2281408

Validated by the manufacturer by flow cytometry on C57BL/6 mouse bone marrow cells, 6 citations

APC anti-mouse CD202b (Tie-2, CD202): BioLegend, #124010, clone TEK4
 RRID: AB_10897106

Validated by the manufacturer by flow cytometry on bEnd.3 mouse endothelial cells, 3 citations

PE anti-mouse CD202b (Tie-2, CD202): BioLegend, #124008, clone TEK4

RRID: AB_2287338

Validated by the manufacturer by flow cytometry on bEnd.3 mouse endothelial cells, 5 citations

APC anti-mouse CD105: BioLegend, #120414, clone MJ7/18

RRID: AB_2277914

Validated by the manufacturer by flow cytometry on bEnd.3 mouse endothelial cells, 5 citations

Pacific Blue anti-mouse CD105: BioLegend, #120412, clone MJ7/18

RRID: AB_209889

Validated by the manufacturer by flow cytometry on bEnd.3 mouse endothelial cells, 2 citations

APC anti-mouse CD90.2: BioLegend, #105312, clone 30-H12

RRID: AB_313183

Validated by the manufacturer by flow cytometry on C57BL/6 mouse thymocytes, 12 citations

FITC anti-mouse Ly-51 (6C3): BioLegend, #108305, clone 6C3

RRID: AB_313362

Validated by the manufacturer by flow cytometry on C57BL/6 mouse bone marrow cells, 6 citations

PerCP-eFluor 710 anti-mouse CD200: eBioscience, #46-5200-82, clone OX90

RRID: AB_10598213

Validated by the manufacturer by flow cytometry on 57BL/6 splenocytes, 1 citation

Biotin anti-mouse CD51: BD Pharmingen, #551380, clone RMV-7

RRID: AB_394174

Validated by the manufacturer by flow cytometry on BALB/c bone marrow leukocytes, 8 citations

PE anti-Ki67: BD Pharmingen, #556027, clone B56

RRID: AB_2266296

Validated by the manufacturer by flow cytometry on permeabilized MOLT-4 cells, 14 citations

PE anti-active Caspase 3: BD Pharmingen, #550821, clone C92-605

RRID: AB_393906

Validated by the manufacturer by flow cytometry on camptothecin treated Jurkat cells, 3 citations

AlexaFluor 647 anti-SOX9: Cell Signaling Technology, #71273, clone D8G8H

RRID: AB_2799799

Validated by the manufacturer by flow cytometry on HeLa cells (blue) and A-204 cells, 12 citations

APC anti-mouse CD146: BioLegend, #134712, clone ME-9F1

RRID: AB_2563088

Validated by the manufacturer by flow cytometry on Mouse endothelial cells, 3 citations

Pacific Blue anti-mouse F4/80: BioLegend, #123124, clone BM8

RRID: AB_893475

Validated by the manufacturer by flow cytometry on Thioglycolate-elicited Balb/c mouse peritoneal macrophages, 21 citations

PerCP/Cy5.5 anti-mouse CD31: BioLegend, #102522, clone MEC13.3

RRID: AB_2566761

Validated by the manufacturer by flow cytometry on C57BL/6 mouse splenocytes, 11 citations

Primary antibodies for immunohistochemical analysis:

BrdU: Bio-Rad, #OBT0030, clone BU 1/75-ICR1, Rat

RRID: AB_609568

Validated by the manufacturer for immunohistochemistry on formalin-fixed paraffin-embedded tissue, 31 citations

CD31: BD Biosciences, 550274, clone MEC 13.3, Rat

RRID: AB_393571

Validated by the manufacturer for immunohistochemistry on zinc-fixed paraffin-embedded section of U-87 MG tumor in mouse brain, 7 citations

type I collagen: Novus Biologicals, clone NB600-408, polyclonal, rabbit

RRID: AB_343276

Validated by the manufacturer for immunohistochemistry on FFPE sections of mouse pancreas tissue and rat colon tissue, 28 citations

type II collagen: Merck Millipore, MAB8887, monoclonal, mouse, clone 6B3

RRID: AB_2260779

Validated by the manufacturer for immunohistochemistry on fetal cartilage, 30 citations

SOX9: Novus Biologicals, NBP1-85551, polyclonal, rabbit

RRID: AB_11002706

Validated by the manufacturer for immunohistochemistry on FFPE sections of human colorectal cancer, glioma, skeletal muscle and small intestine, 4 citations

CPT1a: Cell Signaling Technology, #12252, clone D3B3, rabbit
 RRID: AB_2797857
 Validated by the manufacturer on HeLa, PANC-1 and MCF7 cells, 16 citations

GLUT1: Cell Signaling Technology, #12939, clone D3J3A, rabbit
 RRID: AB_2687899
 Validated by the manufacturer on HepG2 and Huh6 cells, 9 citations

FoxO3a: Cell Signaling Technology, #2497, clone 75D8, rabbit
 RRID: AB_836876
 Validated by the manufacturer on SH-SY5Y cells IGF-I or LY294002 treated, 230 citations

Primary antibodies for immunocytochemical analysis:

FoxO1: Cell Signaling Technology, #2880, clone C29H4, rabbit
 RRID: AB_2106495
 Validated by the manufacturer for immunofluorescent analysis in IGROV-1 cells, 409 citations

FoxO3a: Cell Signaling Technology, #2497, clone 75D8, rabbit
 RRID: AB_836876
 Validated by the manufacturer for immunofluorescent analysis on SH-SY5Y cells, 230 citations

Primary antibodies for Western Blot:

SOX9: Novus Biologicals, NBP1-85551, polyclonal, rabbit
 RRID: AB_11002706
 Validated by the manufacturer for western blot analysis mammalian HEK293T cells, mouse NIH-3T3 cells and rat NBT-II cells, 2 citations

FoxO1: Cell Signaling Technology, #2880, clone C29H4, rabbit
 RRID: AB_2106495
 Validated by the manufacturer for western blot analysis on extracts from IGROV-1 and COS-7 cells, 409 citations

FoxO3a: Cell Signaling Technology, #2497, clone 75D8, rabbit
 RRID: AB_836876
 Validated by the manufacturer for western blot analysis on extracts from Jurkat and PC3 cells, 230 citations

LC3B: Cell Signaling Technology, #3868, clone D11, rabbit
 RRID: AB_2137707
 Validated by the manufacturer for western blot analysis on extracts of various cell line treated with chloroquine, 31 citations

β -actin: Sigma-Aldrich, #A5441, clone AC-15, mouse
 RRID: AB_476744
 Validated by the manufacturer for western blot analysis on cultured human or chicken fibroblast cell extracts, 299 citations

Lamin A/C: Santa Cruz Biotechnology, #sc-376248, clone E-1, mouse
 RRID: AB_10991536
 Validated by the manufacturer for western blot analysis on cell extracts of different mouse and human cell lines, 4 citations

Primary antibodies for ChIP-qPCR:

FoxO1: Abcam, #ab39670, rabbit
 RRID: AB_732421
 Validated by the manufacturer for ChIP analysis on mouse T cells, 33 citations

FoxO3a: Abcam, #ab12162, rabbit
 RRID: AB_298893
 Validated by the manufacturer for ChIP analysis on pig coronary artery endothelial cells, 30 citations

β -actin: Sigma-Aldrich, #A5441, clone AC-15, mouse
 RRID: AB_476744
 Validated by the manufacturer for western blot analysis on cultured human or chicken fibroblast cell extracts, 299 citations

Lamin A/C: Santa Cruz Biotechnology, #sc-376248, clone E-1, mouse
 RRID: AB_10991536
 Validated by the manufacturer for western blot analysis on cell extracts of different mouse and human cell lines, 4 citations

Primary antibodies for ChIP-qPCR:
 FoxO1: Abcam, #ab39670, rabbit
 RRID: AB_732421
 Validated by the manufacturer for ChIP analysis on mouse T cells, 33 citations

FoxO3a: Abcam, #ab12162, rabbit

RRID: AB_298893

Validated by the manufacturer for ChIP analysis on pig coronary artery endothelial cells, 30 citations

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	C3H10T1/2 cells were obtained from the RIKEN Cell Bank
Authentication	None of the cell lines used were authenticated
Mycoplasma contamination	Cell lines were routinely tested for mycoplasma contamination and found negative
Commonly misidentified lines (See ICLAC register)	The C3H10T1/2 cell line is not among the commonly misidentified cell lines

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Analysis was performed on 3-5 day-old mice or 8-10 week-old male and female mice. C57BL/6J mice (Janvier Labs), 129/Sv mice (Janvier Labs), B6.Cg-Tg(CAG-EGFP) mice (Hadjantonakis, A. K. et al., Mech. Dev., 1998), B6.Cg-Tg(Col1a1-cre/ERT2,-DsRed)1Smkm/J mice (Ouyang, Z. et al., Bone, 2014), B6;129S4-Sox9tm1.1Tlu/J mice and B6.Cg-Tg(CAG-DsRed*MST)1Nagy/J mice (The Jackson Laboratory) were used in this study. All colonies were housed and bred in individually ventilated cages in the animal facility of the KU Leuven.
Wild animals	The study did not use any wild animals
Field-collected samples	The study did not include field-collected samples
Ethics oversight	All animal experiments were conducted according to the regulations and with approval of the Animal Ethics Committee of the KU Leuven.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Isolation of mouse skeletal stem cells was adapted from a previously described protocol (Chan, C. K., et al., Cell, 2015). Long bones of 3-5 day old mice were dissected, muscle was cleared away and bones were minced using a scalpel. Bone fragments were then digested in α -MEM supplemented with 3mg/ml collagenase II, 4mg/ml dispase (both from Gibco) and 100U/ml DNase I (Sigma) at 37°C for 45 minutes, with replacement of the digest medium every 15 minutes. Cell suspensions were passed through a 70 μ m cell strainer, washed with PBS containing 2% FBS and stained with antibodies against CD45, Ter119, Tie2, CD105, CD90.2, 6C3 (BioLegend), CD51 (BD Pharmingen) and CD200 (eBioscience), and with the viability dye 7-aminoactinomycin D (BD Pharmingen). For the isolation of skeletal muscle-derived cell populations, hindlimb skeletal muscles, including quadriceps, soleus, gastrocnemius and tibialis anterior, were dissected from 8-week old CAG-DsRed mice, minced using a scalpel and digested in α -MEM medium supplemented with 3mg/ml collagenase II, 4mg/ml dispase and 100U/ml DNase I at 37°C for 60 minutes. Every 15 minutes samples were pipetted up and down vigorously using a 10ml serological pipette to break up tissue fragments. Cell suspensions were passed through a 70 μ m nylon mesh, washed with PBS containing 2% FBS and stained with antibodies against CD45, Ter119, CD31, F4/80 and CD146 (BioLegend), and with 7AAD (BD Pharmingen).
Instrument	BD LSRII
Software	BD FACSDiva
Cell population abundance	Post-sort purity was not determined
Gating strategy	Phenotypic skeletal stem cells (7AAD-CD45-Ter119-Tie2-CD51+CD105-CD90.2-6C3-CD200+) were sorted on a BD FACSAria II (BD

Gating strategy

Biosciences). Previously-defined gating strategies (Chan, C. K., et al., Cell, 2015) were followed and are given in Extended Data Fig. 9a)

Immunophenotypically-defined macrophages (7AAD-CD45+F4/80+), endothelial cells (7AAD-CD45-Ter119-F4/80-CD31+CD146+) and pericytes (7AAD-CD45-Ter119-F4/80-CD31-CD146+) (Extended Data Fig. 9b) were sorted on a BD FACSARIA II. Gating strategies are described in Extended Data Fig. 9b.

Analysis of SOX9^{high} cells: Gating for SOX9^{high} cells was set to have approximately 10% SOX9^{high} cells in control conditions for all cell types.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Two conserved epigenetic regulators prevent healthy ageing

<https://doi.org/10.1038/s41586-020-2037-y>

Received: 23 April 2018

Accepted: 6 January 2020

Published online: 26 February 2020

 Check for updates

Jie Yuan^{1,3,8}, Si-Yuan Chang^{1,3,8}, Shi-Gang Yin^{2,3,4,8}, Zhi-Yang Liu^{1,3,8}, Xiu Cheng^{2,3,8}, Xi-Juan Liu¹, Qiang Jiang¹, Ge Gao^{1,3}, De-Ying Lin^{1,3}, Xin-Lei Kang¹, Shi-Wei Ye¹, Zheng Chen⁵, Jiang-An Yin¹, Pei Hao², Lubin Jiang^{2,6,8} & Shi-Qing Cai^{1,7,8}

It has long been assumed that lifespan and healthspan correlate strongly, yet the two can be clearly dissociated^{1–6}. Although there has been a global increase in human life expectancy, increasing longevity is rarely accompanied by an extended healthspan^{4,7}. Thus, understanding the origin of healthy behaviours in old people remains an important and challenging task. Here we report a conserved epigenetic mechanism underlying healthy ageing. Through genome-wide RNA-interference-based screening of genes that regulate behavioural deterioration in ageing *Caenorhabditis elegans*, we identify 59 genes as potential modulators of the rate of age-related behavioural deterioration. Among these modulators, we found that a neuronal epigenetic reader, BAZ-2, and a neuronal histone 3 lysine 9 methyltransferase, SET-6, accelerate behavioural deterioration in *C. elegans* by reducing mitochondrial function, repressing the expression of nuclear-encoded mitochondrial proteins. This mechanism is conserved in cultured mouse neurons and human cells. Examination of human databases^{8,9} shows that expression of the human orthologues of these *C. elegans* regulators, BAZ2B and EHMT1, in the frontal cortex increases with age and correlates positively with the progression of Alzheimer's disease. Furthermore, ablation of Baz2b, the mouse orthologue of BAZ-2, attenuates age-dependent body-weight gain and prevents cognitive decline in ageing mice. Thus our genome-wide RNA-interference screen in *C. elegans* has unravelled conserved epigenetic negative regulators of ageing, suggesting possible ways to achieve healthy ageing.

Previous studies have shown that dopamine signalling declines with age^{10,11}, and that higher dopamine levels in aged people improve their cognitive functions¹². In *C. elegans*, an age-related decline in the level of the BAS-1 protein—a shared serotonin (5-HT)- and dopamine-synthesizing enzyme (DOPA decarboxylase; Fig. 1a)—is responsible for the loss of these neurotransmitters and a behavioural deterioration², thus providing a genetically traceable marker of ageing in the nervous system. We therefore performed a genome-wide RNA interference (RNAi) screen for regulators of ageing by examining changes in the level of the BAS-1 protein in *C. elegans*. We used transgenic worms (named *P_{bas-1}::bas-1::gfp*) that express BAS-1 fused to green fluorescent protein (GFP), and individually fed the worms with bacteria expressing different double-stranded RNAs that correspond to roughly 80% of the total predicted genes in *C. elegans* (Fig. 1b). To enhance neuronal uptake of dsRNAs, we introduced SID-1—a channel allowing dsRNA diffusion—into the nervous system of the transgenic worms¹³.

After three rounds of testing, we obtained 59 screening hits that prevented an age-related reduction in the BAS-1 protein level; these genes encode various protein classes, including nucleic-acid-binding proteins, receptors and transporters (Extended Data Fig. 1a–c and Supplementary

Table 1). We then performed gene-network analysis of these screening hits using GeneMANIA software¹⁴ and constructed a coexpression network that reveals the interaction among individual genes and their partners (Extended Data Fig. 1d). In line with the aim of our screening, we found ten hits whose human orthologues are involved in age-related neurodegeneration or cell senescence (Extended Data Fig. 1d and Supplementary Table 1). We next examined whether the genes corresponding to the screening hits affect ageing-related behavioural decline, and found that downregulation of the majority of 20 tested genes improved behavioural performance in pharyngeal pumping in aged worms (Extended Data Fig. 1e). Thus our genome-wide screen for regulators of ageing provides a global view of molecules involved in the ageing nervous system.

BAZ-2, SET-6 and age-related decline

Among the seven most prominent hits, a putative epigenetic reader, BAZ-2, and a putative histone 3 lysine 9 (H3K9) methyltransferase, SET-6, appeared at a key node in the network (Extended Data Fig. 1b, d). They were broadly expressed in the *C. elegans* nervous system, including head, body and tail neurons (Extended Data Fig. 2a), and were colocalized with a

¹Institute of Neuroscience and State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China.

²CAS Key Laboratory of Molecular Virology and Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, China. ³University of Chinese Academy of Sciences, Beijing, China.

⁴Academician (Expert) Workstation of Sichuan Province, The Affiliated Hospital of Southwest Medical University, Luzhou, China. ⁵College of Biological and Environmental Sciences, Zhejiang Wanli University, Ningbo, China. ⁶School of Life Science and Technology, ShanghaiTech University, Shanghai, China. ⁷Shanghai Center for Brain Science and Brain-Inspired

Intelligence Technology, Shanghai, China. ⁸These authors contributed equally: Jie Yuan, Si-Yuan Chang, Shi-Gang Yin, Zhi-Yang Liu, Xiu Cheng. [✉]e-mail: lbjiang@ips.ac.cn; sqcai@ion.ac.cn

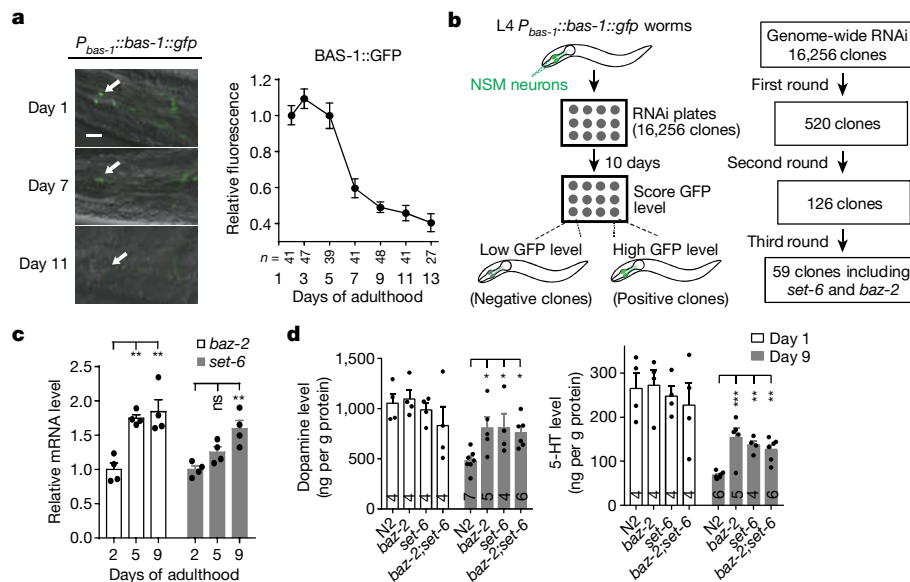


Fig. 1 | A genome-wide screen of *C. elegans* identifies genes that regulate age-related loss of 5-HT and dopamine. **a**, Fluorescent images (left) and quantitative analysis (right) of BAS-1 expression at different ages in transgenic worms. Quantitative analysis of BAS-1 levels was performed by measuring GFP fluorescence intensity in the soma of the NSM neurons. White arrows indicate the NSM neurons. Scale bar, 10 μ m. The numbers of the total tested worms for each data point are shown beneath the x-axis. **b**, Illustration showing the genome-wide RNAi screen. A RNAi library containing 16,256 clones was screened in the first round; RNAi clones that prevented a decline in the BAS-1::GFP level in the transgenic worms at day 9 of adulthood were selected and subjected to further rounds of testing. The clones that markedly increased the

BAS-1::GFP level in ageing transgenic worms in the third round of testing were scored as positive clones. **c**, Age-dependent changes in transcription levels of *baz-2* and *set-6*. $n = 4$ independent experiments. **d**, Dopamine (left) and 5-HT (right) levels in N2, *baz-2*, *set-6* or *baz-2;set-6* mutant worms at day 1 or day 9 of adulthood. The neurotransmitter levels were determined by high-performance liquid chromatography (HPLC). The numbers of independent assays are indicated in each column. All data shown are means \pm s.e.m.; $P < 0.05$; $P < 0.01$; $P < 0.001$; *ns*, not significant (**c** and **d**, one-way analysis of variance (ANOVA) with Dunnett's test; see Supplementary Information for exact *P* values). In **c** and **d**, each data point represents the result of one independent experiment.

nuclear stain, 4',6-diamidino-2-phenylindole (DAPI), indicating that they are nuclear proteins (Extended Data Fig. 2b). Notably, the expression of BAZ-2 and SET-6 increased with age (Fig. 1c and Extended Data Fig. 2c, d).

We then investigated the function of BAZ-2 and SET-6 in modulating the ageing process, and found that deleting *baz-2*, *set-6* or both (double mutant *baz-2;set-6*) increased the levels of BAS-1 (Extended Data Fig. 2e), 5-HT and dopamine (Fig. 1d) in aged worms, but not in young adult worms. We next examined whether the increase in endogenous 5-HT and dopamine caused by deleting *baz-2* or *set-6* could improve behavioural performance in ageing worms. An age-related loss of 5-HT and dopamine causes the decline of many important behaviours in *C. elegans*, including pharyngeal pumping, male mating, and response to food². We found that deleting *baz-2* or *set-6* prevented this age-related behavioural decline, but had no effect on the behaviours of young worms (Fig. 2a–d). This prevention could be reversed by re-expressing *baz-2* or *set-6* through their own promoters in their respective mutant worms (Fig. 2a–d). In support of the notion that epigenetic modulation plays a critical part in modulating longevity^{15,16}, we found that deleting *baz-2* or *set-6* caused a moderate extension of the worm lifespan (Fig. 2e) and enhanced the worm's capacity to resist certain environmental insults (hydrogen peroxide, ultraviolet light and a 35 °C heat shock; Extended Data Fig. 3a–c). Moreover, a loss-of-function mutation in *daf-16*—the key transcription factor in the insulin signalling pathway¹⁷—did not alter the effect of *baz-2* or *set-6* deletion on lifespan extension and resistance to oxidative stress (Extended Data Fig. 3d), suggesting that *daf-16* is not required for the effects of *baz-2* or *set-6* on ageing. By contrast, the effect of deleting *baz-2* or *set-6* was abolished in response to dietary restriction caused by an *eat-2* mutation¹⁸, or in response to reduced mitochondrial function caused by a *clk-1* mutation¹⁹ (Extended Data Fig. 3e, f). Thus, the effect of *baz-2* or *set-6* deletion on the ageing process is likely to be mediated by mechanisms related to dietary restriction and mitochondrial function.

BAZ-2 and SET-6 regulate H3K9 methylation

Notably, we found that the regulation by *baz-2* and *set-6* of age-related behavioural deterioration, longevity and stress response was not additive (Fig. 2a–e and Extended Data Fig. 3a–c), suggesting that *baz-2* and *set-6* act in the same genetic pathway. In support of this notion, we found that endogenous SET-6 and BAZ-2 co-immunoprecipitated (Extended Data Fig. 4a) in homogenates extracted from genome-edited *baz-2::GFP::FLAG*; *set-6::GFP::HA* worms (with *baz-2* tagged by GFP::FLAG and *set-6* tagged by GFP::HA, where HA is haemagglutinin). The BAZ-2 protein belongs to a family of evolutionarily conserved proteins that contain the plant homeodomain (PHD) and the bromodomain, both of which recognize modified histone tails²⁰; SET-6 is a putative H3K9 methyltransferase. To examine whether SET-6 has methyltransferase activity, we performed in vitro histone methylation assays. Incubation of purified truncated proteins containing the SET domain of SET-6 (SET-6_{SET}) with calf histone substrates increased levels of H3K9 dimethylation (H3K9me2) and trimethylation (H3K9me3), but not monomethylation (H3K9me1) (Fig. 3a), indicating that SET-6 is a methyltransferase for H3K9me2 and H3K9me3. Furthermore, we found that deleting *baz-2* or *set-6* reduced the level of global H3K9me3, but not H3K9me1 and H3K9me2 (Extended Data Fig. 4b). Thus, BAZ-2 and SET-6 could function together to regulate the methylation of H3K9 in *C. elegans*.

To investigate which genes are regulated by BAZ-2 and SET-6, we examined the genome-wide distribution of binding sites for these two proteins via chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq) analyses in transgenic worms expressing GFP-fused SET-6 or BAZ-2 (with their own promoters), using anti-GFP antibodies. We found that BAZ-2 and SET-6 co-occupied the promoter region of 2,383 genes (Fig. 3b, Extended Data Fig. 4c, d and Supplementary Table 2), which account for 71.6% and 77.1% of all genes bound by BAZ-2 and SET-6, respectively ($P < 3 \times 10^{-321}$; Fisher's exact test). Among the co-occupied genes, nuclear genes encoding nucleotide-binding

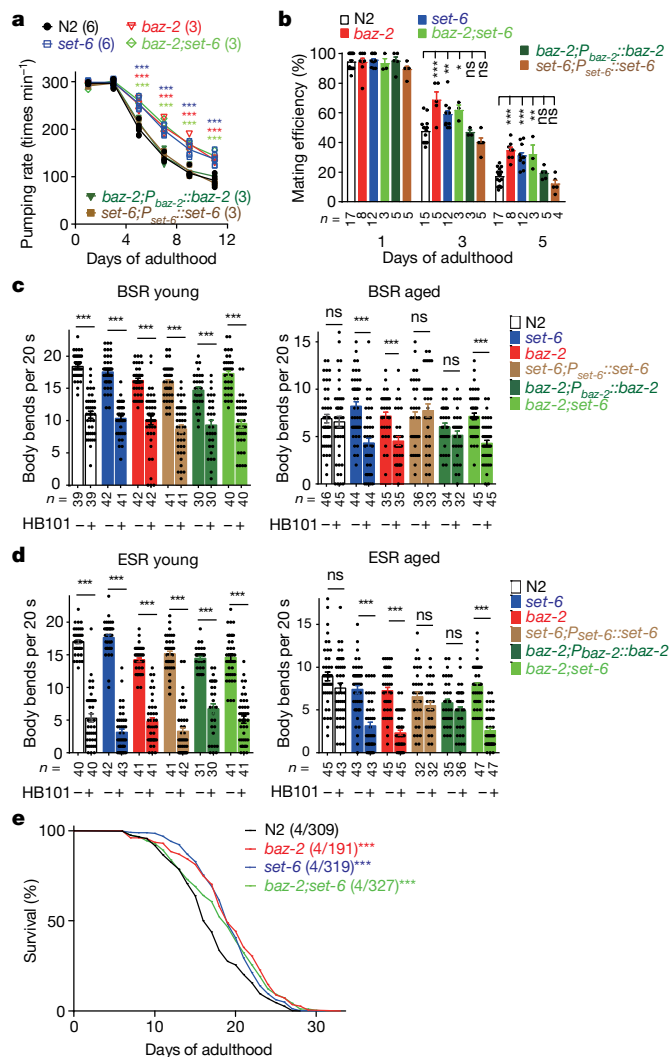


Fig. 2 | Deletion of *baz-2* or *set-6* delays age-related behavioural decline and extends lifespan in *C. elegans*. **a, b**, Age-related decline in pharyngeal pumping (**a**) and male mating (**b**) in worms with different genotypes. The numbers of independent assays are indicated in parentheses (**a**) or are shown beneath the bars (**b**). **c, d**, Basal slowing response (BSR) of well fed worms towards food (**c**), and enhanced slowing response (ESR) of food-deprived worms (**d**), at day 1 (young) or day 9 (aged) of adulthood. BSR and ESR were quantified by the frequency of body bends. The total numbers of tested worms are shown beneath the bars. The *E. coli* strain HB101 was used as the food source in this assay. In **a–d**, data shown are means \pm s.e.m. **e**, Lifespan curves of worms with different genotypes. Data represent the sum of animals in multiple experiments; the numbers of independent assays and tested hermaphrodites are indicated in parentheses. In **a–e**, * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ns, not significant (see Supplementary Information for exact *P* values). **a, b**, One-way ANOVA with Dunnett's test; **c, d**, two-tailed *t*-test; **e**, two-sided log-rank test.

proteins, metal-binding proteins, ribosomal proteins and mitochondrial proteins were enriched (Fig. 3c, d). The occupancy of the endogenously expressed BAZ-2 and SET-6 at the promoter region of those nuclear genes encoding mitochondrial proteins was confirmed by ChIP quantitative polymerase chain reaction (qPCR) analysis in genome-edited *baz-2*^{GFP::FLAG} and *set-6*^{GFP::HA} worms, respectively (Extended Data Fig. 4e, f).

We then investigated transcriptome changes caused by deleting *set-6* or *baz-2*, and found that a total of 450 differentially expressed genes were present in both *baz-2* and *set-6* mutant worms (Fig. 3e, Extended Data Fig. 5a and Supplementary Table 3). These common genes account for 32.3% and 40.3% of the total differentially expressed genes in *baz-2* and *set-6* mutant worms, respectively ($P < 1.3 \times 10^{-126}$; Fisher's exact

test). Notably, ribosomal and mitochondrial proteins were the two most enriched categories among 450 common differentially expressed genes (Fig. 3d), the majority of which were upregulated (Extended Data Fig. 6a). We verified the upregulation of mitochondrial-function-related genes by reverse transcription (RT)–qPCR analysis in both young adult and aged worms (Extended Data Fig. 5b). We next grouped the 450 genes into 5 gene clusters using the *k*-means method, and constructed heat maps of the ChIP–seq data. Interestingly, we found that the binding of BAZ-2 and SET-6 was more prominent at the promoter regions of genes in cluster 2, among which genes related to mitochondrial and ribosomal functions were enriched (Extended Data Fig. 6a), suggesting that BAZ-2 and SET-6 work together to repress the expression of mitochondrial-function-related genes by occupying their promoter regions.

We next determined whether BAZ-2 and SET-6 regulate gene expression by regulating the H3K9 methylation levels of target genes in *C. elegans*. We found that deleting *baz-2* or *set-6* reduced the level of H3K9me3, but not of H3K9me1 and H3K9me2, at cluster 2 genes (Extended Data Fig. 6b–e), suggesting that BAZ-2 and SET-6 regulate the level of H3K9me3 on these genes. Thus, SET-6 and BAZ-2 repress the expression of mitochondrial-function-related genes by regulating the H3K9me3 levels of target genes.

BAZ-2, SET-6 and mitochondrial function

We then inquired whether SET-6 and BAZ-2 regulate mitochondrial functions. We found that deleting and overexpressing *baz-2* or *set-6* markedly elevated and reduced two key mitochondrial activities—ATP production and oxygen consumption—in both young adult and aged worms (Fig. 3f, g). The enhanced mitochondrial activities were not due to increased mitochondrial abundance, because we found that the ratio of mitochondrial DNA to nuclear DNA was unaffected in *baz-2*, *set-6* and *baz-2;set-6* mutant worms (Fig. 3h).

Mitochondrial proteins are encoded by both nuclear and mitochondrial genomes. An imbalance between the expression of proteins from these two sources activates the mitochondrial unfolded protein response (UPR^{mt})^{21,22}, which maintains mitochondrial proteostasis and promotes longevity^{21–25}. Deleting *baz-2* or *set-6* enhanced the expression of a set of nuclear genes encoding mitochondrial proteins, including some mitochondrial ribosomal proteins (Extended Data Fig. 5b). Indeed, we found that an imbalance between mitochondrial versus nuclear proteins (Extended Data Fig. 5c) correlates with UPR^{mt} activation (as revealed by the UPR^{mt} reporter *P_{hsp-6}::GFP*; ref. ²⁶) in *baz-2*, *set-6* and *baz-2;set-6* mutant worms (Fig. 3i, j). This UPR^{mt} activation was abolished by dsRNAs targeting *ubl-5*, a positive regulator of the UPR^{mt} (ref. ²⁷), and by dsRNAs targeting *atfs-1*, a transcription factor mediating the UPR^{mt} (ref. ²⁸; Fig. 3i, j). Furthermore, we found that attenuating UPR^{mt} by downregulating *ubl-5* or *atfs-1* with RNAi prevented the lifespan extension and elevation of pharyngeal pumping ability induced by *baz-2* or *set-6* deletion (Fig. 3k, l). Thus, these two epigenetic factors prevent healthy ageing at least in part via regulating UPR^{mt} activation.

BAZ-2 and SET-6 functions are conserved

The *C. elegans* BAZ-2 has two mammalian homologues, BAZ2A and BAZ2B, as does SET-6, namely EHMT1 and EHMT2 (Extended Data Fig. 7a, b). By analysing the gene-expression profiles of the human prefrontal cortex^{8,9}, we found that the expression of BAZ2B and EHMT1, but not of BAZ2A and EHMT2, increased with age (Fig. 4a and Extended Data Fig. 8a, b). Interestingly, like their *C. elegans* homologues, BAZ2B and EHMT1 co-immunoprecipitated (Extended Data Fig. 8c). Further ChIP–qPCR analysis showed that both BAZ2B and EHMT1 could bind to a set of nuclear genes encoding mitochondrial proteins in HEK293T cells (Extended Data Fig. 8d). In mouse primary neuronal cultures, we found that down-regulating *Baz2b* or *Ehmt1*, or both, enhanced the expression of a set of mitochondrial proteins (Extended Data Fig. 8e and Supplementary

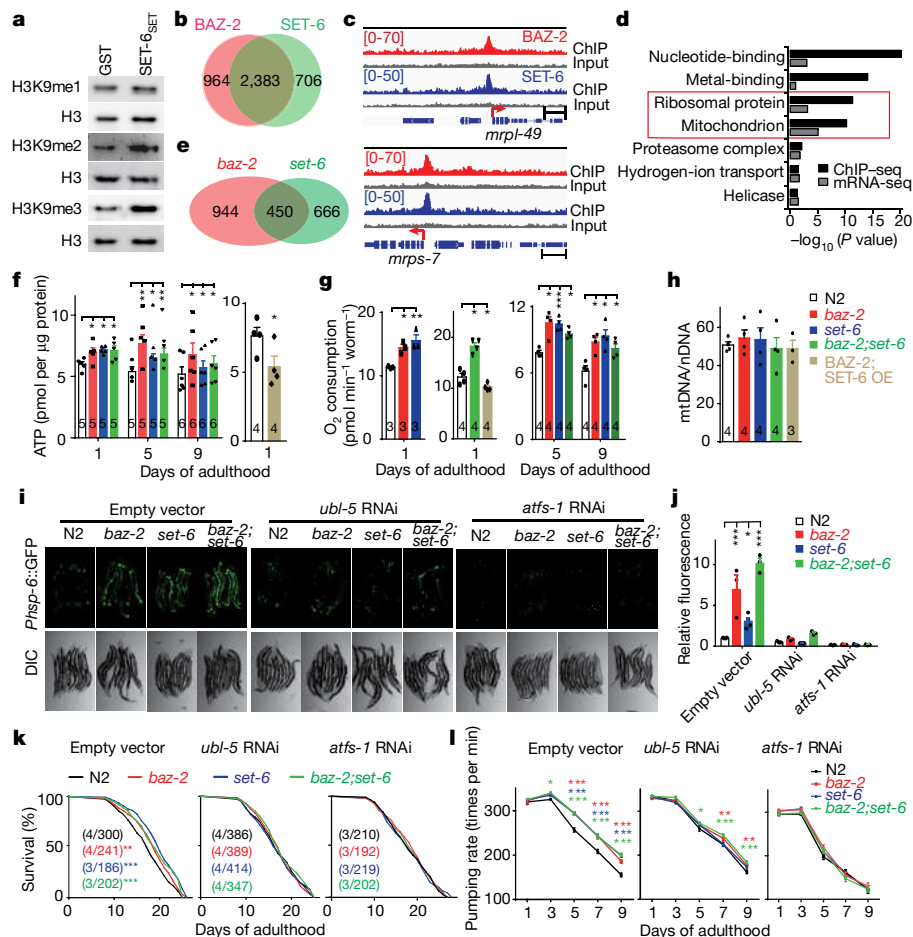


Fig. 3 | BAZ-2 and SET-6 repress mitochondrial function. **a**, The purified glutathione-S-transferase (GST)-fused SET domain of SET-6 proteins (SET-6_{SET}) catalyses the methylation of H3K9me2/3, but not H3K9me1, in vitro. Images are representative of three independent experiments. For gel source data, see Supplementary Fig. 1. **b**, Venn diagram showing the number of genes that bind to BAZ-2 (red) and SET-6 (green) in ChIP-seq data. **c**, ChIP-seq profiles of BAZ-2 or SET-6 binding at two nuclear genes, *mrpl-49* and *mrps-7*, related to mitochondrial function. The ChIP-seq signals are shown in the range [0–70] for BAZ-2 ChIP input and [0–50] for SET-6 ChIP input. Scale bar, 1 kilobase (kb). **d**, Top gene-ontology (GO) terms of 2,383 overlapping BAZ-2- and SET-6-binding genes (black) and 450 overlapping differentially expressed genes in *baz-2* and *set-6* mutant worms (grey). The enrichment of genes was analysed by DAVID Bioinformatics Resources, and the Fisher exact test was used for statistical analyses. **e**, Venn diagram showing differentially expressed genes in mRNA-seq

data from *baz-2* and *set-6* mutant worms. **f–h**, ATP level (**f**), oxygen-consumption rate (**g**), and mitochondrial DNA (mtDNA)/nuclear DNA (nDNA) ratio (**h**) in worms with different genotypes. OE, overexpression. The numbers of independent assays are indicated in the columns. **i, j**, Fluorescent images (**i**) and quantitative analysis (**j**) of worms expressing the UPR^{mt} reporter *P_{hsp-6::GFP}* in the presence of control dsRNAs or dsRNAs targeting *ubl-5* or *atfs-1*. *n* = 3 independent experiments. **k**, Lifespan curves of worms in the presence of control (left), *ubl-5* (middle) or *atfs-1* (right) dsRNAs. Data represent the sum of animals in multiple experiments; the numbers of independent experiments and of tested hermaphrodites are indicated in parentheses; two-sided log-rank test. **l**, Age-dependent decline in pharyngeal pumping in the presence of control (left), *ubl-5* (middle) or *atfs-1* (right) dsRNAs. *n* = 4 independent experiments. In **f–h**, **j**, **l**, data shown are means ± s.e.m.; one-way ANOVA with Dunnett's test. For all assays, **P* < 0.05; ***P* < 0.01; ****P* < 0.001 (see Supplementary Information for exact *P* values).

Table 4). The oxygen consumption and ATP production could also be increased and reduced by downregulation and overexpression, respectively, of *Baz2b* and *Ehmt1* (Extended Data Fig. 8f–i). Thus, *Baz2b* and *Ehmt1* repress mitochondrial functions in mammalian neurons.

Baz2b modulates cognitive ageing in mice

To further explore whether the epigenetic regulator *Baz2b* modulates the behavioural deterioration of ageing mice, we constructed *Baz2b*-null (*Baz2b*^{−/−}) mice by deleting four base pairs in the *Baz2b* gene, which resulted in a frameshift mutation and ablation of the *Baz2b* protein (Extended Data Fig. 9a, b). Consistently, we found that *Baz2b* ablation improved mitochondrial function in the hippocampus and cerebellum of 12-month-old male mice (Extended Data Fig. 9c). Ablation of *Baz2b* also prevented age-dependent weight gain in male mice (Fig. 4b, c), suggesting that it alters energy metabolism in aged mice. Furthermore,

although the null mutation of *Baz2b* did not affect the exploratory and locomotive activities of mice (Extended Data Fig. 9d, e), it did improve Barnes' spatial learning (Fig. 4d) and spatial memory for new locations (Fig. 4e) in old (older than 18-month) male mice. By contrast, there was no apparent difference in these behavioural tests among young (roughly three-month-old) *Baz2b*^{−/−}, *Baz2b*^{+/−} and wild-type male mice (Fig. 4e and Extended Data Fig. 9f), and these male mice showed no difference in lifespan (Fig. 4f). Thus, *Baz2b* contributes to the age-related deterioration of mitochondrial function and cognitive behaviour in mice.

Mitochondrial dysfunction has been implicated in the pathogenesis of Alzheimer's disease—an example of unhealthy ageing of the brain²⁹. We found in an existing human dataset⁹ that *BAZ2B* and *EHMT1* expression in the prefrontal cortex correlate positively with the progression of Alzheimer's disease (Extended Data Fig. 10a, b), and negatively with the expression of key mitochondrial proteins (Extended Data Fig. 10c–h). Given the conserved roles of *BAZ2B* and *EHMT1* in

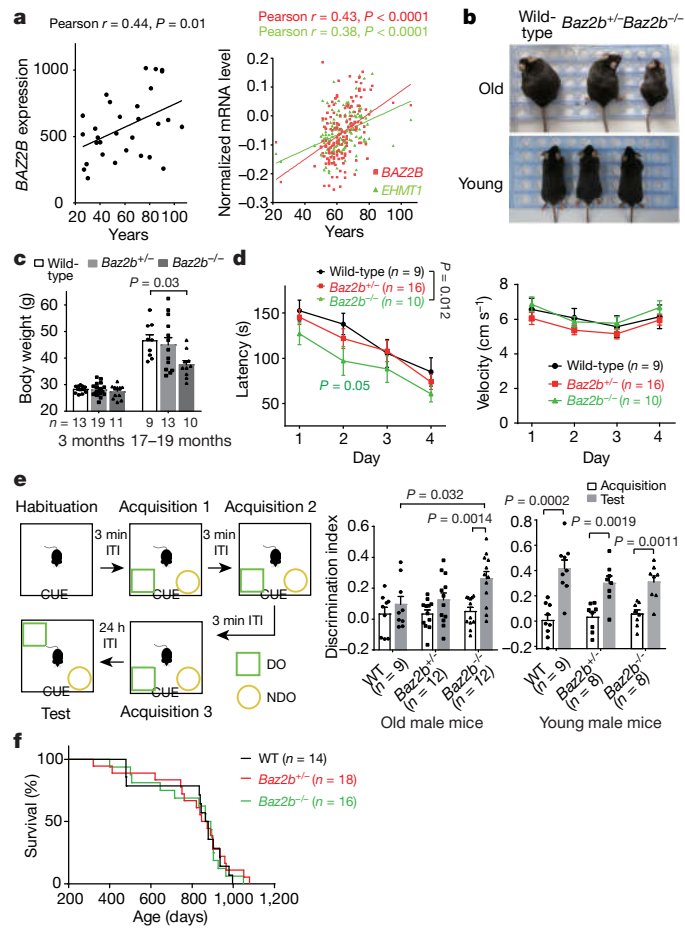


Fig. 4 | Knockout of *Baz2b* improves spatial learning and memory abilities in old mice. **a**, Transcription levels of *BAZ2B* and *EHMT1* in the prefrontal cortex of human brains at different ages. Expression values ($n = 30$ samples) in the left panel are from the dataset GSE1572 (see ref.⁸); expression values ($n = 145$ samples) in the right panel are from brain samples without neurodegenerative diseases in the dataset GSE44772 (see ref.⁹). Pearson's r correlation coefficient was used for statistical testing. **b**, **c**, Representative images of old and young wild-type, $Baz2b^{+/+}$ and $Baz2b^{-/-}$ male mice (**b**) and quantitative analysis of their body weight (**c**). The numbers of tested mice are shown beneath the bars. **d**, Escape latency (left) and locomotive velocity (right) in Barnes maze trials during training days for old wild-type, $Baz2b^{+/+}$ and $Baz2b^{-/-}$ male mice. **e**, Diagram showing the novel-location-recognition (NLR) test (left), and discrimination indexes for old (middle) and young (right) wild-type, $Baz2b^{+/+}$ and $Baz2b^{-/-}$ male mice in the NLR test. Visual cues were on the one side of the training arena wall. DO, displaced object; ITI, intertrial interval; NDO, nondisplaced object. The discrimination index = (time exploring the DO – time exploring the NDO)/(time exploring the DO + time exploring the NDO). In **c**–**e**, data shown are means \pm s.e.m. **f**, Lifespan curves of wild-type, $Baz2b^{+/+}$ and $Baz2b^{-/-}$ male mice. In **d**–**f**, the numbers of tested mice are indicated in parentheses. P values were determined by: **c**, one-way ANOVA with Dunnett's test; **d**, two-way repeated-measures ANOVA with Dunnett's test; **e**, two-tailed t -test; **f**, two-sided log-rank test.

regulating mitochondrial function, their increased expression might contribute to mitochondrial dysfunction in Alzheimer's disease.

Here, by using genome-wide RNAi screening in *C. elegans*, we have provided the first global view of genes that may regulate age-related behavioural deterioration, and identified two repressive epigenetic factors—BAZ-2/BAZ2B and SET-6/EHMT1—that prevent healthy ageing. Notably, ablation of these factors promotes healthy ageing by improving mitochondrial function and cognitive behaviour via the regulation of H3K9 methylation levels at target genes (Extended Data Fig. 10i). These findings suggest that preventing age-related mitochondrial impairment

by targeting repressive epigenetic regulators is a potential strategy for improving behavioural performance and achieving healthy ageing.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2037-y>.

- Guarente, L. Aging research—where do we stand and where are we going? *Cell* **159**, 15–19 (2014).
- Yin, J. A., Liu, X. J., Yuan, J., Jiang, J. & Cai, S. Q. Longevity manipulations differentially affect serotonin/dopamine level and behavioral deterioration in aging *Caenorhabditis elegans*. *J. Neurosci.* **34**, 3947–3958 (2014).
- Bansal, A., Zhu, L. J., Yen, K. & Tissenbaum, H. A. Uncoupling lifespan and healthspan in *Caenorhabditis elegans* longevity mutants. *Proc. Natl Acad. Sci. USA* **112**, E277–E286 (2015).
- Beard, J. R. et al. The world report on ageing and health: a policy framework for healthy ageing. *Lancet* **387**, 2145–2154 (2016).
- Hansen, M. & Kennedy, B. K. Does longer lifespan mean longer healthspan? *Trends Cell Biol.* **26**, 565–568 (2016).
- Yin, J. A. et al. Genetic variation in glia-neuron signalling modulates ageing rate. *Nature* **551**, 198–203 (2017).
- Crimmins, E. M. & Beltrán-Sánchez, H. Mortality and morbidity trends: is there compression of morbidity? *J. Gerontol. B* **66**, 75–86 (2011).
- Lu, T. et al. Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).
- Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
- Hedden, T. & Gabrieli, J. D. Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci.* **5**, 87–96 (2004).
- Bäckman, L., Nyberg, L., Lindenberger, U., Li, S. C. & Farde, L. The correlative triad among aging, dopamine, and cognition: current status and future prospects. *Neurosci. Biobehav. Rev.* **30**, 791–807 (2006).
- Chowdhury, R. et al. Dopamine restores reward prediction errors in old age. *Nat. Neurosci.* **16**, 648–653 (2013).
- Calixto, A., Chelur, D., Topalidou, I., Chen, X. & Chalfie, M. Enhanced neuronal RNAi in *C. elegans* using SID-1. *Nat. Methods* **7**, 554–559 (2010).
- Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C. & Morris, Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* **9** (Suppl 1), S4 (2008).
- Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell Biol.* **16**, 593–610 (2015).
- Sen, P., Shah, P. P., Nativio, R. & Berger, S. L. Epigenetic mechanisms of longevity and aging. *Cell* **166**, 822–839 (2016).
- Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. daf-16: an HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
- McKay, J. P., Raizen, D. M., Gottschalk, A., Schafer, W. R. & Avery, L. eat-2 and eat-18 are required for nicotinic neurotransmission in the *Caenorhabditis elegans* pharynx. *Genetics* **166**, 161–169 (2004).
- Ewbank, J. J. et al. Structural and functional conservation of the *Caenorhabditis elegans* timing gene clk-1. *Science* **275**, 980–983 (1997).
- Santoro, R., Li, J. & Grummt, I. The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat. Genet.* **32**, 393–396 (2002).
- Houtkooper, R. H. et al. Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* **497**, 451–457 (2013).
- Mouchiroud, L. et al. The NAD(+)/sirtuin pathway modulates longevity through activation of mitochondrial UPR and FOXO signaling. *Cell* **154**, 430–441 (2013).
- Durieux, J., Wolff, S. & Dillin, A. The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **144**, 79–91 (2011).
- Merkwirth, C. et al. Two conserved histone demethylases regulate mitochondrial stress-induced longevity. *Cell* **165**, 1209–1223 (2016).
- Tian, Y. et al. Mitochondrial stress induces chromatin reorganization to promote longevity and UPR(mt). *Cell* **165**, 1197–1208 (2016).
- Yoneda, T. et al. Compartment-specific perturbation of protein handling activates genes encoding mitochondrial chaperones. *J. Cell Sci.* **117**, 4055–4066 (2004).
- Benedetti, C., Haynes, C. M., Yang, Y., Harding, H. P. & Ron, D. Ubiquitin-like protein 5 positively regulates chaperone gene expression in the mitochondrial unfolded protein response. *Genetics* **174**, 229–239 (2006).
- Haynes, C. M., Yang, Y., Blais, S. P., Neubert, T. A. & Ron, D. The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Mol. Cell* **37**, 529–540 (2010).
- Lin, M. T. & Beal, M. F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443**, 787–795 (2006).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Worm strains and culture

The wild-type Bristol N2, *set-6(ok2195)*, TU3401 (*sid-1(pk3321);uls69*), *daf-16(mu86)*, *eat-2(ad1116)* and *clk-1(qm30)* strains were obtained from the Caenorhabditis Genetics Center. The *baz-2(tm0235)* strain was obtained from the National Bioresource Project, Japan. The *set-6;baz-2* double mutant worms were generated by crossing *set-6(ok2195)* with *baz-2(tm0235)*. The TU3401;*P_{bas-1}::bas-1::gfp* strain was generated by crossing TU3401 with *P_{bas-1}::bas-1::gfp* transgenic worms (SQC0017). The *baz-2;P_{bas-1}::bas-1::gfp* and *set-6;P_{bas-1}::bas-1::gfp* strains were generated by crossing *baz-2(tm0235)* and *set-6(ok2195)* with *P_{bas-1}::bas-1::gfp* worms, respectively. The *baz-2;daf-16* and *set-6;daf-16* mutant worms were generated by crossing *daf-16(mu86)* with *baz-2(tm0235)* and *set-6(ok2195)* animals, respectively. The *baz-2;eat-2* and *set-6;eat-2* mutant worms were generated by crossing *eat-2(ad1116)* with *baz-2(tm0235)* and *set-6(ok2195)* animals, respectively. The locus of *baz-2* is adjacent to that of *clk-1*, so the *baz-2;clk-1* strain was generated by deleting *baz-2* (*yfh0100* allele) in the genomic background of *clk-1(qm30)* mutant worms using the CRISPR–Cas9 system. The *set-6;clk-1* worms were generated by crossing *set-6(ok2195)* with *clk-1(qm30)*. The *P_{hsp-6}::gfp* transgenic strain SJ4100 was crossed with *set-6(ok2195)*, *baz-2(tm0235)*, and *set-6;baz-2* worms to express *P_{hsp-6}::GFP* in these mutant worms. All worms were cultivated at 20 °C on nematode growth medium (NGM) plates seeded with *Escherichia coli* OP50 unless stated otherwise.

Mice

The null mutation in the *Baz2b* gene of C57BL/6J mice was generated using the CRISPR–Cas9 system by the Suzhou Non-human Primate Facility, Institute of Neuroscience, Chinese Academy of Sciences. The single-guide RNAs (sgRNAs) targeting exon 2 of the *Baz2b* gene were microinjected into the cytoplasm of C57BL/6J embryos with *Cas9* messenger RNA. The injected embryos were transferred into the oviducts of pseudopregnant mice to produce *Baz2b* mutant mice. The mutation in the *Baz2b* gene was confirmed by PCR and DNA sequencing. The sgRNA sequence was 5'-gaaactgctgaagccacgga-3'. After backcrossing to wild-type C57BL/6J mice for three generations, *Baz2b*-null mice were identified and used. Mice used for experiments were littermates from crosses between *Baz2b* heterozygotes. All mice were housed under specific pathogen-free conditions on a 12-h light/dark cycle (lights were on from 07:00 to 19:00 every day).

Molecular biology

The *baz-2* gene with its 3,769-base-pair (bp) upstream promoter was amplified from N2 genomic DNA and inserted into the pPD95.75 vector to obtain the *P_{baz-2}::baz-2::gfp* plasmid. The *set-6* complementary DNA was amplified from N2 transcripts. The *set-6* cDNA and 981-bp *set-6* promoter were inserted into the pPD95.75 vector to generate the *P_{set-6}::set-6::gfp* plasmid. To construct *P_{baz-2}::baz-2::3 × FLAG* and *P_{set-6}::set-6::3 × HA* plasmids, we fused 3 × FLAG and 3 × HA sequences to the termini of *baz-2* DNA and *set-6* cDNA, respectively. The fosmid (catalogue number 00754076101505996 E05) expressing GFP-fused BAZ-2 was obtained from TransgeneOme³⁰. To express SET-6_{SET} proteins (from the 453th amino acid to the 708th amino acid of SET-6), we amplified the *set-6_{set}* cDNA and inserted it into the pGEX-4T-1 vector between the BamHI and EcoRI restriction-enzyme sites.

To express genes in mammalian cells, we inserted *BAZ2B* and *EHMT1* cDNAs into pCI-neo and pcDNA3.1(+) vectors, respectively. Mouse *Baz2b* and *Ehmt1* cDNAs were inserted into pKH3 and c-Myc-pcs2+MT vectors, respectively. The small hairpin RNA (shRNA) sequences targeting *Baz2b* (5'-ggctcttctccaagtaa-3') and *Ehmt1* (5'-gaggatagtagcattcta-3')³¹ or a nonsense negative control (NC) shRNA sequence (5'-ttctccgaacgtgtcacgt-3') were inserted into pLKD vectors between the AgeI and EcoRI restriction-enzyme sites, respectively.

All plasmids were verified by DNA sequencing. We used the following primers to construct the plasmids: *baz-2* promoter forward, 5'-aacctgcaggaagtcctgcgacgacaag-3' and reverse, 5'-tccccccgggttttgggaagaattacatg-3'; *baz-2* DNA forward, 5'-tccccccgggatgagtgataactcatatcatg-3' and reverse, 5'-ctcagacgccttcacgttcaccgggtgaca-3'; *set-6* promoter forward, 5'-acatgcacgcctcttttagaataatacaac-3' and reverse, 5'-acgcgtcgactttctataagcagtaaac-3'; *set-6* cDNA forward, 5'-cgcggatccatggaacgatctcgaactgg-3' and reverse, 5'-ccgctcagtgatcttcgtcggcaggttc-3'; *BAZ2B* cDNA forward, 5'-atggagtcgtgagaacgggttacc-3' and reverse, 5'-tcagctcactttgaaagtatctgtcc-3'; *EHMT1* cDNA forward, 5'-atggccgcccgcgatccga-3' and reverse, 5'-tcatagggggctcggcggcagc-3'; *Baz2b* cDNA forward, 5'-atggagtcgtgagaactgttg-3' and reverse, 5'-tcagctcactttgaaagtatctg-3'; *Ehmt1* cDNA forward, 5'-ctcatttctgaaggacttgaattcaatggccgctgatgctga-3' and reverse, 5'-acgactcactatgttctagatcataggggtcagcagcgg-3'.

Transgenic worms and CRISPR–Cas9 genome editing

All transgenic worms were generated following the standard protocol³² by injecting the respective plasmids. Following is detailed information on transgenic strains: SQC0508 *yfhx0508* (*P_{baz-2}::baz-2::gfp* fosmid (obtained from <http://transgeneome.mpi-cbg.de/>; ref. ³⁰) at 50 ng μl⁻¹; *rol6(su1006)* at 10 ng μl⁻¹); SQC0505 *yfhEx0505* (*P_{set-6}::set-6::gfp* at 50 ng μl⁻¹; *P_{lin-44}::mCherry* at 10 ng μl⁻¹); SQC0509 *yfhx0509* (*P_{baz-2}::baz-2::gfp* at 50 ng μl⁻¹; *P_{lin-44}::mCherry* at 10 ng μl⁻¹; TM0235); SQC0510 *yfhx0510* (*P_{set-6}::set-6::gfp* at 50 ng μl⁻¹; *P_{lin-44}::mCherry* at 10 ng μl⁻¹; VC2683); SQC0519 *yfhx0519* (*P_{baz-2}::baz-2::3 × FLAG* at 50 ng μl⁻¹; *P_{set-6}::set-6::3 × HA* at 50 ng μl⁻¹; *P_{lin-44}::gfp* at 10 ng μl⁻¹); SQC0520 *yfhx0520* (*P_{set-6}::set-6::3 × HA* at 50 ng μl⁻¹; *P_{lin-44}::gfp* at 10 ng μl⁻¹).

CRISPR–Cas9-mediated genome editing was performed as described⁶. To construct *baz-2^{GFP::FLAG}* and *set-6^{GFP::HA}* genome-edited worms, we inserted GFP::FLAG and GFP::HA tags into the 3'-ends of the *baz-2* and *set-6* genes, respectively, by Cas9-triggered homologous recombination. The sgRNA sequences were inserted into the pDD162 vector (Addgene). A homologous-repair template with a flexible linker, the *gfp* coding sequence and the 3 × FLAG (or 3 × HA) coding sequence were cloned into the pPD95.75 vector as the donor plasmid. The sgRNA (50 ng μl⁻¹), donor plasmids (10 ng μl⁻¹), and 2.5 ng μl⁻¹ pCFJ90 (Addgene) were co-injected into young adult N2 hermaphrodites and tag insertions were confirmed by PCR and DNA sequencing. The sgRNAs targeting *baz-2* were 5'-cagaaaaagtaaccggtta-3' and 5'-aatcatccattgatgttc-3'; the sgRNAs targeting *set-6* were 5'-tggcagttatgaattctcgt-3' and 5'-tgcaacttcattgctgaactt-3'. To construct *baz-2(yfh0100)* mutant worms, sgRNAs targeting *baz-2* were inserted into the pDD162 vector. The sgRNAs were 5'-ggaacatcagcatcaacgt-3' and 5'-tttgattatatacatcaaa-3'. The *baz-2(yfh0100)* mutant has a 1,026-bp deletion and a 165-bp insertion. The sequences deleted in *baz-2(yfh0100)* covered part of the first exon and the whole second exon (from the 156th to the 1,181th base of the *baz-2* genomic sequence). The insertion site was found at the 155th base of the *baz-2* genomic sequence.

RNAi screening

To enhance the neuronal uptake of dsRNAs, we generated TU3401;*P_{bas-1}::bas-1::gfp* worms by crossing *P_{bas-1}::bas-1::gfp* transgenic worms with TU3401 worms, which express a dsRNA channel, SID-1, in neurons¹³. We performed a genome-wide RNAi screen using TU3401;*P_{bas-1}::bas-1::gfp* worms. The RNAi clones from the Ahringer *C. elegans* library (Source BioScience LifeScience) were cultured in liquid broth with 100 μg ml⁻¹ carbenicillin overnight and then seeded on NGM plates containing 25 μg ml⁻¹ carbenicillin, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and 20 μM 2'-deoxy-5-fluorouridine (FUDR). Bacteria carrying the empty vector L4440 were used as the control. Approximately 30 synchronized *P_{bas-1}::bas-1::gfp*;TU3401 transgenic worms were transferred to RNAi agar plates at the L4 stage. The GFP intensity of NSM neurons in the transgenic worms was examined ten days later. We selected those clones that were estimated to preserve 'high' BAS-1::GFP levels

Article

in more than 50% of tested aged worms for re-examination in a second round of testing. 'High' means that the level of GFP fluorescence in aged worms is comparable with that in young adult worms. Only one plate with about 30 worms was analysed in the first round of screening, and the fluorescence was estimated by a single investigator. In the second round of testing, the clones obtained from the first round of testing were re-examined independently by two investigators for three biological repeats with the same clones. The percentage of worms with 'high' GFP fluorescence was recorded, and RNAi clones with a mean score (average of replicates) of more than 50% were selected and re-examined in the third round of screening by measuring GFP fluorescence using a confocal microscope. RNAi clones that markedly increased BAS-1 levels in the third round of testing were regarded as positive clones.

HPLC

The dopamine and serotonin levels in *C. elegans* were detected using HPLC as described². Briefly, age-synchronized worms were cultured on NGM plates and harvested at day 1 or day 9 of adulthood in M9 buffer. Then, 100 μ l of packed worms were sonicated in 0.3 M perchloric acid (containing 2 mM EDTA \cdot 2Na) and centrifuged at 17,000g for 15 min at 4 °C. Samples were filtered through a 0.22- μ m filter (Millipore), and then a 25- μ l sample was used for HPLC detection. Quantification of dopamine and 5-HT was carried out by comparing peak areas and retention times with the respective peak characteristics of commercial standards (Sigma, USA). The protein level of each sample was quantified by the Bradford assay. Dopamine and 5-HT levels were normalized to the protein level of each sample.

C. elegans behavioural assays

Pharyngeal pumping rates of synchronized worms were scored on cultivating plates. We examined the number of pharyngeal contractions in 10 s using a dissection microscope as described⁶.

Male mating assays were conducted as described². About 40 synchronized L4 males were picked out and cultured on a 60-mm plate. Two males of each genotype and two young adult N2 hermaphrodites were picked into a 35-mm NGM plate seeded with OP50 lawn and allowed to mate for 24 h; the males were then removed. Successful mating was defined as emergence of more than three male progeny in a mating plate. The mating efficiency was calculated as the percentage of plates with successful mating in total mating plates.

We measured the basal slowing response (BSR) and enhanced slowing response (ESR) of worms as described³³. We incubated 60-mm NGM plates seeded with a ring of *E. coli* HB101 at 37 °C for 3 h, and then allowed the plates to cool down to room temperature. Simultaneously, we prepared NGM plates without food. For ESR, worms were deprived of food for 30 min. Body bends within 20 s were scored under a dissection microscope.

The lifespan of worms was examined as described, with some modifications⁶. A total of roughly 90 synchronized worms were seeded on three OP50 plates at 20 °C. We assessed the number of dead worms and picked them out daily.

We carried out RNAi of the *ubl-5* and *atfs-1* genes as described⁶. Activation of UPR^{mt} in young adult worms was reported by $P_{hsp-6}::GFP$ fluorescence. Pharyngeal pumping and lifespan were examined using plates seeded with *ubl-5*, *atfs-1* or control dsRNAs throughout the experiment. We examined the pharyngeal pumping of worms treated with RNAi for the genes identified in the screening hits in the presence of 20 μ M FUDR. The empty vector L4440 was used as a control.

Stress assays

Hydrogen peroxide stress assay. About 20 synchronized worms at day 5 of adulthood were transferred to each well, which contained 800 μ l worm S-basal buffer with various concentrations of H₂O₂ in a 12-well plate at 20 °C. Four hours later, 200 units of catalase (Sigma, C9322) were added to neutralize the H₂O₂, and the mortality of worms was scored.

Ultraviolet stress assay. About 30 synchronized hermaphrodites at day 5 of adulthood were exposed to 1,600 J m⁻² ultraviolet light in an NGM plate without bacteria. After recovery, worms were cultured in standard condition, and the mortality of ultraviolet-treated worms was scored daily.

Heat-shock assay. About 30 synchronized hermaphrodites at day 5 of adulthood were incubated at 35 °C, and their mortality was checked every 2 h.

Immunoprecipitation and western blotting

For co-immunoprecipitation of BAZ-2 and SET-6, mixed stages of transgenic worms or genome-edited worms were harvested and washed for three times with M9 buffer. Samples were frozen in liquid nitrogen, and ground to fine powder using a pestle and mortar. The worm lysate was sonicated in lysis buffer containing 50 mM HEPES pH 7.4, 1 mM EGTA, 1 mM MgCl₂, 150 mM KCl, 10% glycerol, 1 mM NaF, 1 mM phenylmethylsulfonyl fluoride (PMSF), and complete protease-inhibitor cocktail (Roche 11836145001). After treatment with NP-40 at a final concentration of 0.05%, samples were lysed at 4 °C for 1 h and then centrifuged at 20,000g for 15 min at 4 °C to remove the debris. The supernatant was incubated with anti-FLAG antibodies (conjugated to agarose beads) overnight at 4 °C. After incubation, the agarose beads were washed with lysis buffer and boiled in SDS sample buffer (2% SDS, 10% glycerol, 50 mM Tris-HCl, pH 6.8, 0.01% bromophenol blue, 5% β -mercaptoethanol) for 10 min. Samples were separated on SDS-PAGE gels and the standard western blot procedure was carried out. The primary antibodies used were anti-FLAG (Abmart, M2008) and anti-HA (Roche, 11867431001).

For co-immunoprecipitation of BAZ2B and EHMT1, HEK293T cells stably expressing BAZ2B-FLAG were transfected with 3 \times HA-EHMT1 plasmids. Cells were lysed with ice-cold NE10 buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1 mM MgCl₂, 0.1% Triton X-100) and centrifuged at 1,000g for 10 min at 4 °C; nuclear proteins were then extracted with NE150 buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1 mM MgCl₂, 0.1% Triton X-100, 150 mM NaCl). Lysate was cleared by centrifugation at 20,000g for 20 min at 4 °C. The supernatant was diluted with the same volume of dilution buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 20% glycerol, 0.1% Triton X-100, 0.4 mM EDTA) and then incubated with anti-FLAG antibodies conjugated to agarose beads for 4 h at 4 °C for further analysis.

For immunostaining of H3K9 methylation, worms at synchronized day 2 of adulthood, cultured on NGM plates containing 20 μ M FUDR, were harvested and washed several times in M9 buffer, then frozen in liquid nitrogen. Worm samples were boiled in SDS sample buffer for 15 min, separated on SDS-PAGE gels, and analysed by standard western blotting. Primary antibodies used were anti-H3K9me1 (Abcam ab9045), anti-H3K9me2 (Abcam ab115159), anti-H3K9me3 (Abcam ab8898) and anti-H3 (Abcam ab1791). Second antibodies were IRDye 800CW goat anti-rabbit IgG (LI-COR P/N 925-32211) and IRDye 680RD goat anti-rabbit IgG (LI-COR P/N 925-68071). The stained membranes were scanned using the Odyssey CLx imaging system to measure fluorescence at 800 nm and 700 nm. The normalized H3K9 methylation levels were calculated by normalizing the ratio of H3K9 methylation and histone 3 levels to that of N2 worms.

In vitro histone methyltransferase assay

We incubated 1–2 μ g of GST-fused SET-6_{SET} proteins with 1–4 μ g of histone proteins (LS002544, Worthington) and S-adenosyl-L-(methyl-3H) methionine (SAM; sigma) in a mixture of 20 μ l methylase activity buffer (50 mM Tris-HCl, pH 8.0, 10% glycerol, 20 mM KCl, 5 mM MgCl₂, 1 mM dithiothreitol (DTT) and 1 mM PMSF) for 16 h at 20 °C. We then added SDS loading buffer, and boiled the samples for 15 min. Proteins were resolved on a 15% SDS-PAGE gel and visualized by western blotting.

Chromatin immunoprecipitation

For ChIP assays, chromatin immunoprecipitation was performed as described³⁴ with some modifications. Briefly, young adult or aged

worms were harvested and washed several times in M9 buffer, and then washed twice in phosphate-buffered saline (PBS). Worms were lysed in crosslinking buffer (1% formaldehyde in PBS with proteinase-inhibitor cocktail) using a glass dounce homogenizer on ice, and fixed on a Nuta-tor shaker for 15 min at 37 °C. Fixed worm samples were quenched with 0.125 M glycine and washed in cold PBS buffer with protease inhibitors three times. Then samples were resuspended in ice-cold FA buffer (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% sodium deoxycholate, 1% Triton X-100, 1 mM PMSF and protease-inhibitor cocktail) with 0.1% SDS, and sonicated using a Bioruptor sonication system (Diagenode UCD-200) at high amplitude for 10–15 cycles of 30 s on and 30 s off. The sonicated samples were centrifuged at 16,000g for 15 min at 4 °C. Then, samples were precleaned with protein A/G agarose beads and immunoprecipitated overnight using GFP-trap agarose beads (Chromotec ACT-CM-GFA0250) at 4 °C. After proteinase K digestion and reverse crosslinking, the precipitated DNA and input DNA were purified using phenol/chloroform, precipitated with ethanol and subjected to DNA library construction or ChIP–qPCR assays.

For ChIP–seq analysis of histone H3K9 methylation, chromatin was digested using 6 µl of micrococcal nuclease (MNase) (CST catalogue number 10011S) in 400 µl of buffer B (CST catalogue number 7007) containing 0.5 mM DTT for 20 min at 37 °C, and then was stopped with 25 mM EDTA. About 10 µg of the digested chromatin was used for immunoprecipitation, and 5% of the extract was saved as the input sample. The antibodies used were H3K9me1 (Abcam ab9045), H3K9me2 (Abcam ab1220) and H3K9me3 (Abcam ab8898).

For ChIP–seq library preparation, 20 ng purified DNA samples were used. Library construction was performed using the QIAseq Ultralow Input Library Kite (QIAGEN 180495) according to the manufacturer's recommendations. DNA libraries were sequenced on an Illumina HiSeq X-ten instrument, with 150-bp paired-end sequencing. Three batches of worms were collected for each ChIP–seq measurement.

For ChIP–qPCR analysis of genes bound to BAZ2B or EHMT1 in HEK293T cells, immunoprecipitation was performed using anti-IgG, anti-FLAG (Sigma, F3165) or anti-EHMT1 (R&D SYSTEMS, PP-B042200) antibodies. Immunoprecipitated DNA was purified using phenol/chloroform extraction and then used for qPCR analysis. Primer sequences for ChIP–qPCR are listed in the Supplementary Information.

ChIP–seq data analysis

Raw reads were filtered using cutadapt (version 1.15)³⁵ to obtain clean reads with the following parameters: -q 20,20 -m 18 -a -A -o -p. Clean reads were aligned to the reference worm genome (WormBase, <https://wormbase.org>, version WS266) using Bowtie2 aligner (version 2.3.3.1)³⁶ with default parameters. ChIP–seq peaks were called using MACS software (version 2.1.0.20150731)³⁷. For BAZ-2 and SET-6 ChIP–seq datasets, MACS2 was used to call the narrow peaks with the default parameters. All called peaks were annotated using the HOMER (version 4.9, 2-20-2017)³⁸ annotatePeaks.pl function with custom annotation mode (using the WS266 genome sequence and a GTF file). The aligned sequence-alignment map (SAM) files were converted into binary alignment map (BAM) files and then indexed using samtools software (version 1.5)³⁹ with 'samtools view -@ 10 -bS -q 30' and 'samtools index -@ 10' parameters. To obtain read signal tracks for the Integrative Genomics Viewer (IGV, Broad Institute), we converted indexed BAM files into bigwig files using the reads per kilobase of transcript per million mapped reads (RPKM) normalization method with deeptools (version 2.5.4)⁴⁰, using 'bamCoverage -b o -normalizeUsingRPKM -numberOfProcessors=6 -extendReads 200 -binSize=10' parameters. Bedtools (version 2.26.0)⁴¹ with 'bedtools intersect -a -b -wa -wb -e -f 0.5 -F 0.5' parameters were used to obtain the overlapping peaks of BAZ-2 and SET-6. Genes with enriched ChIP–seq reads were used to carry out functional-enrichment analysis on the DAVID website (<https://david.ncifcrf.gov>, version 6.8)⁴², using the functional module. Enriched items were represented as a bar plot using the seaborn package (version 0.8.1) of Python. An in-house

Python scripts integrated by the metaseq (version 0.5.5.4)⁴³ framework was used to generate heatmap plots of the differentially expressed genes (obtained by RNA sequencing). The differentially expressed genes were listed in the same order for heatmaps of both mRNA-seq data and BAZ-2/SET-6 ChIP–seq data.

RNA sequencing

Worms were synchronously cultured on NGM plates, and transferred to NGM plates containing 20 µM FUDR. Worms were harvested at day 2 of adulthood, washed several times using M9 buffer, and then lysed with TRIzol reagent (Invitrogen). Total RNA was extracted using the RNeasy Mini Kit (QIAGEN). RNA quality was assessed with the Agilent 2100 bioanalyser system, and samples with an RNA intensity (RIN) above 8.0 were used to construct the library. Then, mRNAs were purified with oligo(dT) magnetic beads and cDNAs were synthesized using oligo(dT) primers. Bead-bound cDNAs were digested with the restriction enzyme NlaIII and ligated with the Illumina adaptor 1 at the sticky 5'-end. Bead-bound cDNA fragments were then digested with MmeI and ligated with Illumina adaptor 2 at the 3'-ends. After linear PCR amplification, products were purified by PAGE electrophoresis. Sequencing was performed with an Illumina HiSeq 2000 sequencer. Two batches of worms were collected for mRNA sequencing.

RNA-seq analysis

Differentially expressed genes were defined with the following criteria: upregulated genes (false discovery rate (FDR) less than 0.001, log₂-transformed fold change greater than 1, transcripts per million (TPM)_{set-6 or baz-2} greater than 5); downregulated genes (FDR less than 0.001, log₂-transformed fold change less than -1, and TPM_{N2} greater than 5); and other genes representing the ones that are not differentially expressed. We used the seaborn packages of Python to plot scatter figures. Differentially expressed genes of each sample (*baz-2* and *set-6* mutant worms) were screened out and then their log₂-transformed fold change values were collected together to generate a pandas DataFrame for further use. Heatmaps with five *k*-means clusters were generated using the seaborn clustermap function. Each gene cluster from the heatmap was used to analyse GeneOntology (GO) function enrichment in the DAVID website.

qPCR for mRNA and mtDNA quantification

We collected synchronized young adult or aged worms and washed off bacteria with M9 buffer. Total RNAs were extracted and cDNAs were generated using the QuantiTect reverse transcription kit (Qiagen, catalogue number 205314). RT–PCR reactions were conducted using the SYBR Premix Ex Taq kit (Takara, RR420A) and a Light Cycler 480 (Roche). To quantify mitochondrial DNA (mtDNA), we lysed 20 young adult worms in 40 µl worm PCR lysis buffer (50 mM KCl, 10 mM Tris pH 8.3, 2.5 mM MgCl₂, 0.45% NP-40, 0.45% Tween-20, 0.01% gelatin, with freshly added 0.1 mg ml⁻¹ proteinase K). DNAs were released by heating the worms in a PCR cycler at 65 °C for 90 min and 95 °C for 15 min. The worm lysate was diluted 50 times with nuclease-free water and 5 µl of the sample was used as the template for RT–PCR. The ratio of values for a mitochondrial gene, *nd-1*, and a nuclear gene, *act-3*, is used as the relative level of mtDNA per nuclear genome. Primers used for RT–PCR are listed in Supplementary information.

Measurement of ATP and oxygen consumption

To quantify worm ATP levels, we collected about 150 young adult or aged worms, washed them three times to remove bacteria, and resuspended them in 100 µl M9 buffer. After five freeze/thaw cycles (from liquid nitrogen to 40 °C water), worm pellets were boiled for 25 min. Samples were cooled on ice and centrifuged at 11,000g for 10 min at 4 °C. The supernatant was carefully transferred to a new tube. The samples were diluted four times with double-distilled water and then subjected to ATP measurement. The results were normalized to the protein level of each sample.

Article

To measure the oxygen-consumption rates (OCRs) in *C. elegans*, we rinsed off about 200 synchronized young adult and aged worms and washed them three times with M9 buffer by gravity separation. The worm pellets were divided into five replicated tubes containing 1 ml M9 buffer. After standing for 30 min, the samples were pipetted into wells of Seahorse assay plates, and 1 ml M9 buffer was placed into the blank well. Oxygen consumption was measured ten times at 22 °C with the following protocol: 1, calibrate probes; 2, loop 10 times; 3, mix 2 min; 4, time delay 2 min; 5, measure 2 min; 6, loop end. The OCR value was normalized to the number of worms per well.

To measure ATP levels in mammalian cells, we lysed primary cultured neurons in 10 mM Tris-HCl (pH 7.6), 1 mM EDTA and 0.5% Triton X-100 buffer for 20 min at 4 °C. Samples were then centrifuged at 20,000g for 10 min at 4 °C, and the supernatants were diluted 20 times with lysis buffer. ATP levels of the samples were measured using a bioluminescence detection kit (Promega G7570). The ATP level was normalized to the protein level of each sample. OCRs of primary neurons were measured using the Seahorse XF24 analyzer (Agilent). The culture medium was changed to XF base medium containing 10 mM glucose, 2 mM glutamax, 1 mM sodium pyruvate and 2% B27, pH 7.4, and then cells were cultured in a 37 °C incubator without CO₂ for 1 h. After incubation, OCRs were measured and the following respiratory poisons were sequentially added to the assay plate at final concentrations of: oligomycin A, 1 μM (Selleckchem S1478); FCCP, 4 μM (Sigma C2920); rotenone, 1 μM (Sigma R8875). The results were normalized to the total protein amount in each well.

To measure mitochondrial OCRs, we isolated mitochondria from different brain regions of *Baz2b*^{-/-}, *Baz2b*^{+/-} and wild-type male mice as described⁴⁴. Mitochondrial OCRs were measured using a Seahorse XF24 analyser. We plated 10 μg mitochondria in 50 μl mitochondrial assay solution (MAS) buffer (70 mM sucrose, 220 mM mannitol, 5 mM KH₂PO₄, 5 mM MgCl₂, 2 mM HEPES, 1 mM EGTA, 0.2% fatty-acid BSA, pH 7.2) in each well of an XF24 plate, and added MAS buffer without mitochondria in four wells for background correction. We centrifuged mitochondria at 2,000g at 4 °C for 10 min and then added 450 μl MAS buffer containing substrates (5 mM succinate, 5 mM malate, 5 mM glutamate, 5 mM pyruvate) to each well. Mitochondrial OCRs were measured by sequentially adding 0.6 mM ADP, 18 μM oligomycin, 9 μM FCCP and 5 μM rotenone together with 2 μM antimycin to the assay plate.

Primary neuron culture

Primary cortical and cerebellar neurons were prepared from cerebral and cerebellar tissues of mice at embryonic day 14.5 (E14.5) and postnatal day 6 (P6), respectively, as described^{45,46}. *Baz2b* and *Ehmt1* plasmids were electroporated into primary cortical neurons using Nucleofector (Lonza) in the presence of P3 primary cell 4D-Nucleofector solution (Lonza, V4XP-3024). The transfected neurons were seeded into plastic culture dishes, coated with poly-D-lysine (Sigma, P7280) and laminin (Invitrogen 23017015), at a density of 200,000 cells per square centimetre. The neurons were maintained in serum-free neurobasal medium (Invitrogen, 21103049) supplemented with B-27 (Invitrogen, 17504044), 2 mM glutamax-1 (Invitrogen, 35050061), and penicillin/streptomycin (Hyclone, SV30010). The electroporated neurons were assayed at day in vitro (DIV) 2 or 3. For shRNA-mediated downregulation of *Baz2b* and *Ehmt1*, primary cerebellar neurons at DIV3 were infected with lentivirus carrying *Baz2b* or *Ehmt1* shRNA sequence and cultured in serum-free NeuroBasal-A medium (Thermo, 10888022), 2% B27 supplements, 1% glutamax-1, 1% penicillin/streptomycin and 250 μM KCl. Transfected neurons were harvested at DIV8 for further experiments.

Mouse behavioural assays

Young (3-month-old) and old (more than 18-month-old) male *Baz2b*^{-/-}, *Baz2b*^{+/-} mice and wild-type littermates were used in all experiments. The ages of old *Baz2b*^{-/-}, *Baz2b*^{+/-} and wild-type mice were 21.1 ± 3.3, 21.2 ± 2.9 and 21.1 ± 3.4 months, respectively. Animals were handled once a day for one week before behavioural tests. All procedures were

approved by the Animal Care and Use Committee of the Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China.

Open field test

The open field test was performed as described⁴⁷. Mice were placed in the centre of a polystyrene box (40 cm × 40 cm × 40 cm) and the behavioural activity of each mouse was recorded for 15 min using the EthoVision video tracking system (Noldus, Wageningen, The Netherlands).

Barnes maze test

A modified Barnes maze test was performed as described⁴⁸, using an opaque polystyrene disc of 120 cm in diameter. The maze contained 40 holes, and a black polystyrene escape box was placed under one of these holes. Distinct visual cues around the maze were used throughout the study, and an overhead light was used as an aversive stimulus. Mice were trained for four consecutive days. Four trials at intervals of 15 min were performed each day. At the beginning of training on the first day, mice were gently guided by the experimenter to the escape hole and were then covered with a black box for 2 min. In each training trial, mice were given 3 min to find the escape hole. If they found the escape hole within 3 min, the exact time they spent in finding the hole was scored as the escape latency. If they failed to find the escape hole within 3 min, they were gently guided towards the hole and the allotted investigation time was regarded as the escape latency. Before each trial, the escape box and the maze were cleaned with 75% alcohol.

Novel-location recognition

A novel-location-recognition test was performed as described⁴⁹. Mice were placed in the training arena (24 cm × 24 cm × 24 cm) for one 10-min session without objects and for another three 10-min sessions with two distinct objects. During the inter-session interval of 3 min, mice were returned to their home cages. The objects used were a conical flask (height 8 cm; depth 4 cm) and a toy brick (8 cm × 4 cm × 4 cm). Before each session, the arena and the objects were cleaned with 75% alcohol. Twenty-four hours after training, mice were tested for 10 min in the original arena, in which one of objects (displaced object, DO) was displaced to a new place and the other object (nondisplaced object, NDO) was not moved. During training and testing, exploration was recorded with a digital camera and scored using software described in ref.⁵⁰ by an experimenter who was blind to genotypes. Exploration was defined as sniffing or touching the objects with the nose (but not climbing on, turning around or biting objects). The discrimination index is calculated as follows: (time exploring the DO minus time exploring the NDO), divided by (time exploring the DO plus time exploring the NDO).

Lifespan assay

An experienced technician checked the health condition of mice daily and estimated whether a humane endpoint had been reached. A mouse had reached a humane endpoint when it showed more than one of the following conditions: (1) serious trauma; (2) no eating and drinking for more than 24 h; (3) no response to gentle disturbance that lasted for a long time; and (4) a rapid loss of more than 20% body weight. Some mice were found dead in their home cage at the daily inspection.

Data reporting and statistics

No statistical methods were used to predetermine sample size. The sample sizes in our experiments were determined from related published analyses. The experiments were not randomized. All *C. elegans* strains were synchronized and cultivated at 20 °C. The RNAi screen and the behavioural, lifespan and stress assays were performed at room temperature (roughly 20 °C). Behavioural and lifespan experiments were repeated at least three times and investigators were blinded to the genotypes or dsRNA treatments. Worms were picked at L4 stage and raised to a range of ages, chosen in an unbiased manner for behavioural, lifespan and imaging assays. In lifespan and male mating assays,

worms that crawled to the wall of the plate were not included in the data. HEK293T cell lines (catalogue number SCSP502) were ordered from the cell bank of the Chinese Academy of Sciences. All procedures for culturing mouse primary neurons and performing mouse behavioural tests were approved by the Animal Care and Use Committee of the Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China. The investigators were blinded to genotypes during mouse behavioural assays. For the novel-location recognition test, mice that did not explore for more than 3 seconds in total for both objects during training or testing were excluded from the analysis. Animals that had discrimination indexes of more than 0.2 or less than -0.2 during training were considered to have a significant location/object bias during training and were also excluded from further analysis.

We used GraphPad Prism 7 (GraphPad Software, Inc.) for statistical analyses. We tested the normality of the data with the Shapiro–Wilk normality test. We used the Brown–Forsythe test to examine differences in variance between groups. We used: two-tailed Student's *t*-test to analyse differences between two groups; one-way analysis of variance (ANOVA) followed Dunnett's correction test to analyse differences between multiple groups; two-way ANOVA to analyse differences between multiple groups with two variations; and a two-sided log-rank (Mantel–Cox) test to analyse lifespan statistics. The variance of all plots and graphs is represented as means \pm s.e.m.; *n* refers to the number of worms, mice or independent experiments. The significance of statistical differences is indicated as: **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

We used two publicly available microarray datasets, GSE1572 (ref.⁸) and GSE44772 (ref.⁹), to analyse changes in gene expression between normal and pathological ageing of human brains. All brain samples in the dataset GSE1572 were used for analysis. In Fig. 4a (right panel) and Extended Data Fig. 8b, control brain samples without neurodegenerative disease in the dataset GSE44772 are regarded as normal ageing brains; liquid-nitrogen-preserved samples with RNA intensity numbers (RINs) of 5 or more are used for correlation analysis. We used samples from ageing brains with Alzheimer's disease from dataset GSE44772 for the correlation analysis shown in Extended Data Fig. 10a–h; we removed samples that had RINs of less than 5, were from people younger than 50, were affected by Huntington's disease, or had been preserved on dry ice. Pearson's *r* correlation was performed using GraphPad Prism 7.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The raw sequence data generated here have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under accession number PRJNA554977. Source Data for Figs. 1–4 and Extended Data Figs. 1–10 are provided with the paper.

Code availability

All custom code used to generate figures is available at <https://github.com/SHYKON-YIN/nature>.

30. Sarov, M. et al. A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*. *Cell* **150**, 855–866 (2012).
31. Ohno, H., Shinoda, K., Ohyama, K., Sharp, L.Z. & Kajimura, S. EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. *Nature* **504**, 163–167 (2013).
32. Hope, I.A. (ed.) *C. Elegans: A Practical Approach* (Oxford Univ. Press, 1999).
33. Sawin, E. R., Ranganathan, R. & Horvitz, H. R. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* **26**, 619–631 (2000).
34. Mukhopadhyay, A., Deplancke, B., Walhout, A. J. & Tissenbaum, H. A. Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to

- study transcription factor binding to DNA in *Caenorhabditis elegans*. *Nat. Protocols* **3**, 698–709 (2008).
35. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
36. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
37. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
38. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
39. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
40. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
41. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
42. Huang, da W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).
43. Dale, R. K., Matzat, L. H. & Lei, E. P. metaseq: a Python package for integrative genome-wide analysis reveals relationships between chromatin insulators and associated nuclear mRNA. *Nucleic Acids Res.* **42**, 9158–9170 (2014).
44. Frezza, C., Cipolat, S. & Scorrano, L. Organelle isolation: functional mitochondria from mouse liver, muscle and cultured fibroblasts. *Nat. Protocols* **2**, 287–295 (2007).
45. Viesselmann, C., Ballweg, J., Lumbard, D. & Dent, E. W. Nucleofection and primary culture of embryonic mouse hippocampal and cortical neurons. *J. Vis. Exp.* **47**, 2373 (2011).
46. Lee, H. Y., Greene, L. A., Mason, C. A. & Manzini, M. C. Isolation and culture of post-natal mouse cerebellar granule neuron progenitor cells and neurons. *J. Vis. Exp.* **23**, 990 (2009).
47. Patel, J. C., Rossignol, E., Rice, M. E. & Machold, R. P. Opposing regulation of dopaminergic activity and exploratory motor behavior by forebrain and brainstem cholinergic circuits. *Nat. Commun.* **3**, 1172 (2012).
48. Patil, S. S., Sunyer, B., Höger, H. & Lubec, G. Evaluation of spatial memory of C57BL/6J and CD1 mice in the Barnes maze, the multiple T-maze and in the Morris water maze. *Behav. Brain Res.* **198**, 58–68 (2009).
49. Wimmer, M. E., Hernandez, P. J., Blackwell, J. & Abel, T. Aging impairs hippocampus-dependent long-term memory for object location in mice. *Neurobiol. Aging* **33**, 2220–2224 (2012).
50. Xu, X. et al. Modular genetic control of sexually dimorphic behaviors. *Cell* **148**, 596–607 (2012).

Acknowledgements We thank M.-m. Poo and X. Shi for critical reading of the manuscript; Q. Sun, Y.-J. Cai and Y.-Z. Li for generating Bax2b-knockout mice; the *C. elegans* Genetics Center (funded by the National Institutes of Health (NIH) Office of Research Infrastructure Programs P40 OD010440) for providing strains; members of C.-y. Li's laboratory for helping with mouse behavioural assays; M. Sun for helping to prepare the H3K9 ChIP-seq DNA libraries; and the Institute of Neuroscience Molecular and Cellular Biology Core Facility, the Optical Imaging Facility, and the Animal Facility (Chinese Academy of Sciences) for technical support. This work was supported by grants to S.-Q.C. (from the National Natural Science Foundation of China, grants 31925022, 91949206 and 81527901; and from the National Key R&D Program of China, grant 2018YFC2000400); the Strategic Priority Research Program of the Chinese Academy of Science (grant XDB32020100); the Shanghai Municipal Science and Technology Major Project (grant 2018SHZDZX05); and grants to L.J. (from the National Key R&D Program of China, grant 2018YFA0507300; and the National Natural Science Foundation of China, grant 31771455).

Author contributions J.Y. conducted HPLC assays, imaging assays and most of the behavioural assays; performed co-immunoprecipitation assays and western blots to detect the H3K9 methylation level in *C. elegans*; prepared the mRNA samples for RNA-seq; conducted experiments in mammalian cells; and analysed the gene-expression profile from human brain gene-expression datasets. J.Y. and X.-J.L. performed the genome-wide RNAi screen as well as lifespan and stress-resistance assays. S.-Y.C. conducted the RT-qPCR assays in *C. elegans* and the qPCR-based determination of mitochondrial DNA/nuclear DNA ratios; detected the ATP level and OCR in *C. elegans*; conducted UPR^{mt} activation assays; and performed behavioural and lifespan assays related to UPR^{mt}. S.-Y.C. and X.-J.L. performed the lifespan and H₂O₂ stress resistance assays in mutants of different longevity pathways. X.-J.L. carried out assays of male mating behaviour. S.-G.Y. analysed the ChIP-seq data and the RNA-seq data of *C. elegans*, and helped to analyse gene-expression levels from human brain datasets. X.C. conducted the in vitro histone methyltransferase assay and ChIP-seq assays in *C. elegans*. Z.-Y.L. performed the co-immunoprecipitation and ChIP-qPCR assays in mammalian cells. Z.-Y.L. and S.-W.Y. measured the OCRs of isolated mitochondria from mice. G.G. carried out co-immunoprecipitation of SET-6 and BAZ-2. X.-L.K. and Z.C. helped to prepare samples for ChIP-seq assays in *C. elegans*. J.-A.Y. performed part of the assays for ATP detection in mouse neurons. Z.-Y.L. and D.-Y.L. performed mouse behavioural assays. Q.J. helped with the BSR and ESR assays. P.H. supervised the ChIP-seq data analysis. L.J. supervised experiments regarding epigenetics and analysis of the data. S.-Q.C. conceived the project. S.-Q.C., L. J., J.Y., S.-Y.C. and S.-G.Y. designed the study, analysed the data, and wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

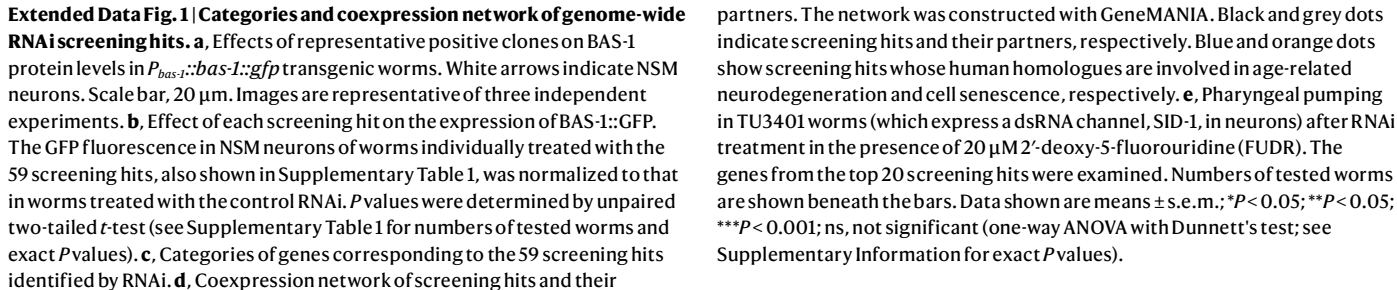
Additional information

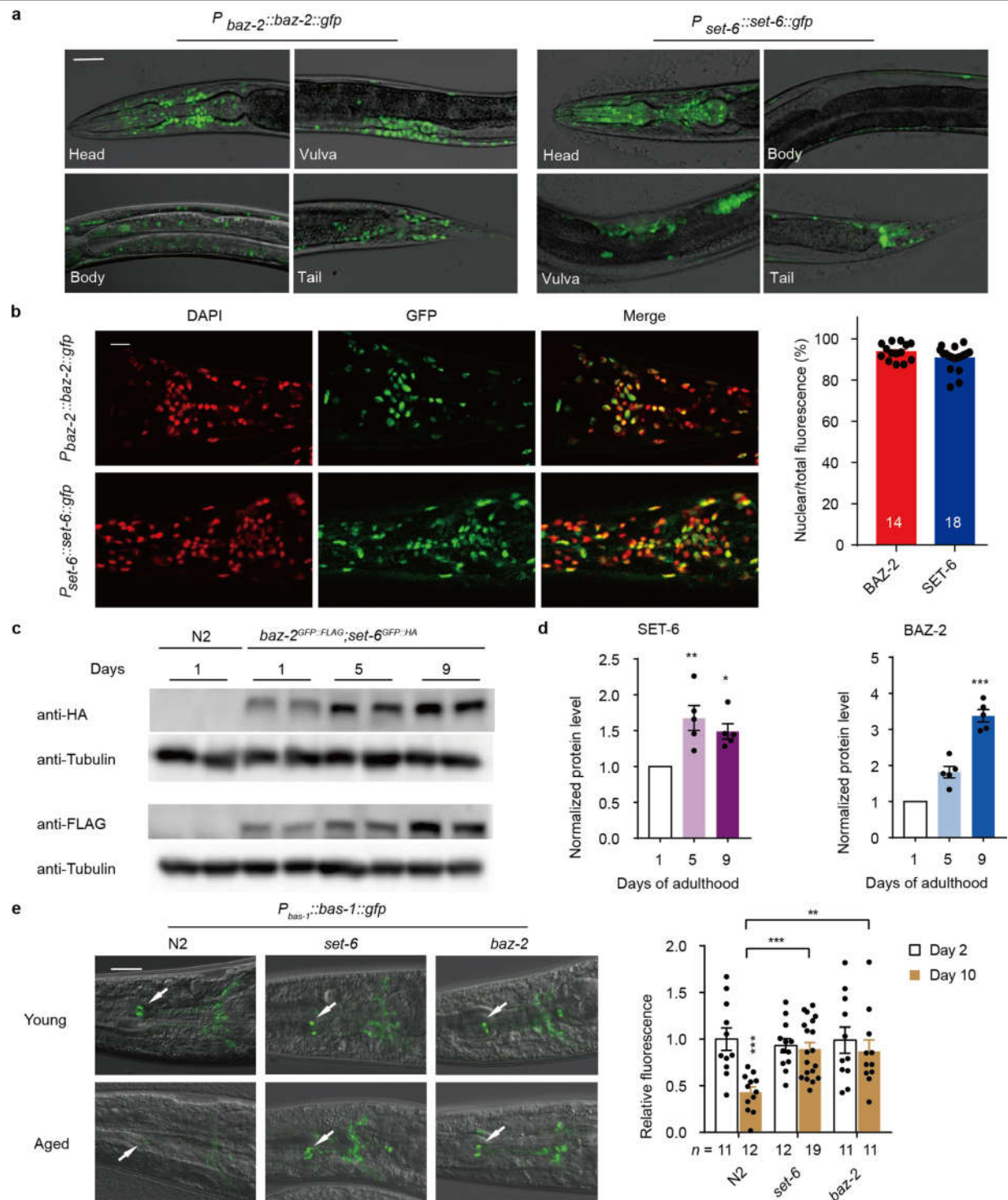
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2037-y>.

Correspondence and requests for materials should be addressed to L.J. or S.-Q.C.

Peer review information Nature thanks Cole Haynes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

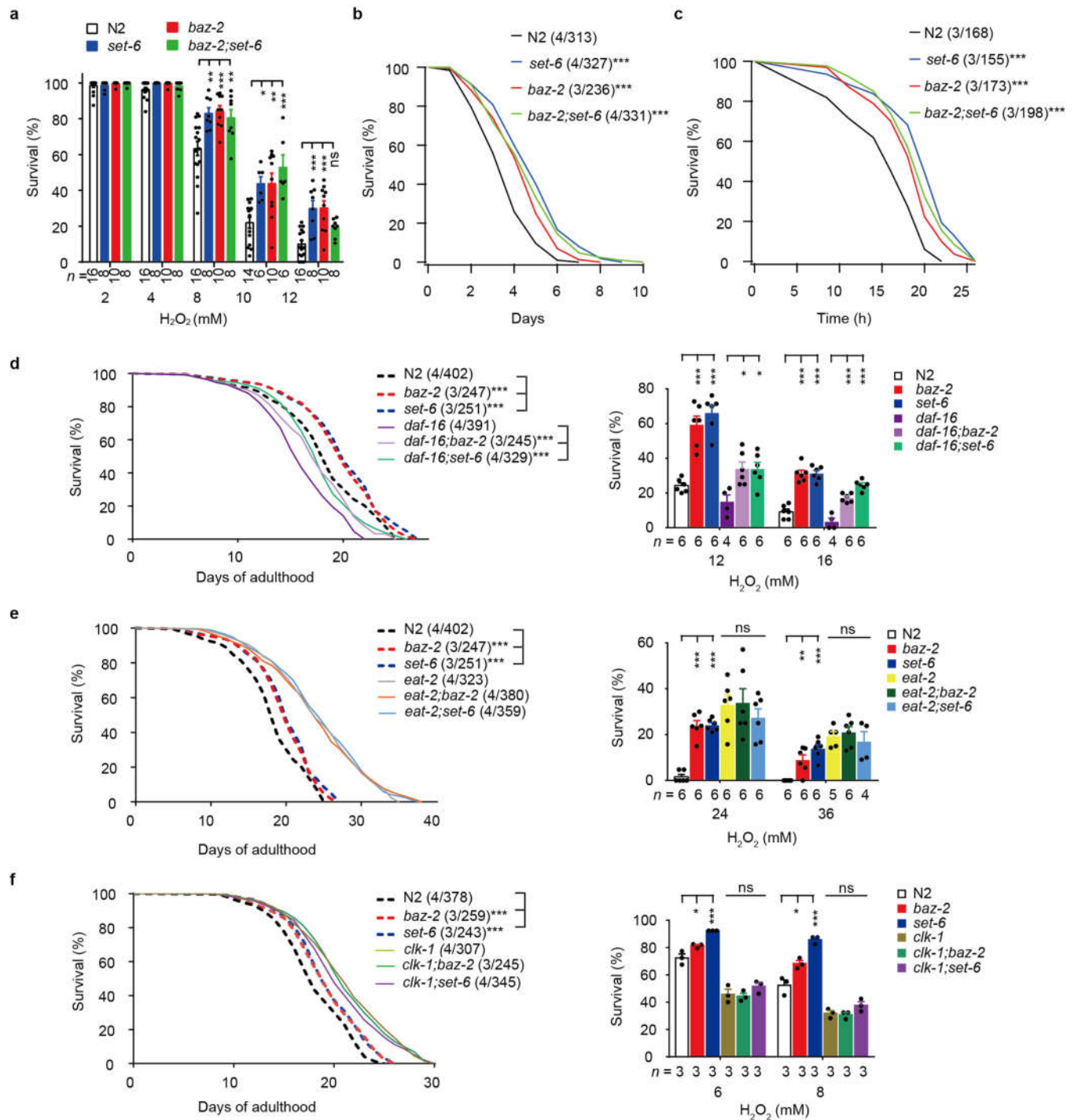




Extended Data Fig. 2 | BAZ-2 and SET-6 regulate age-related decline in BAS-1 expression. **a**, Expression patterns of BAZ-2 and SET-6. Scale bar, 30 μ m. Images are representative of three independent experiments.

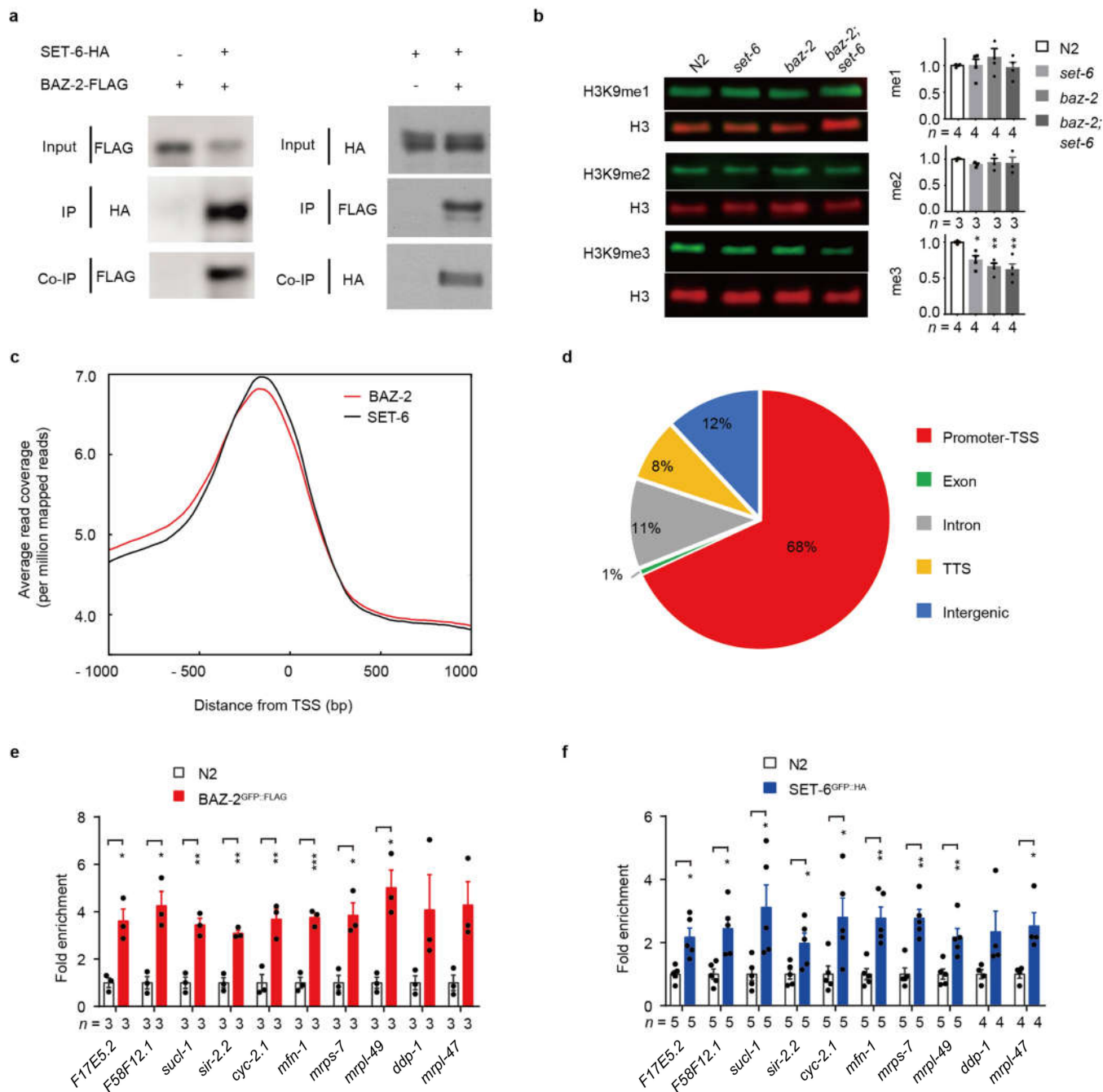
b, Representative fluorescence images (left) and quantitative analysis (right) of BAZ-2 and SET-6 expression. We quantified only those cells whose nuclear morphology was clearly visualized with DAPI. Scale bar, 10 μ m. $n = 14$ and 18 worms for BAZ-2 and SET-6, respectively. **c**, **d**, Representative western blots (**c**) and quantitative analysis (**d**) of age-related changes in endogenous SET-6 and BAZ-2 protein levels in genome-edited *baz-2^{GFP::FLAG};set-6^{GFP::HA}* worms. $n = 5$ independent experiments. Tubulin expression is used as a reference. For gel

source data, see Supplementary Fig. 2. In **d**, each data point represents the result of one independent experiment. **e**, Fluorescence images (left) and quantitative analysis (right) of BAS-1 expression in *baz-2* and *set-6* mutant worms. Quantitative analysis of BAS-1 levels was performed by measuring GFP fluorescence intensity in the soma of NSM neurons. White arrows indicate NSM neurons. Scale bar, 15 μ m. The numbers of tested worms are shown beneath the bars. All data shown are means \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (**d**, Kruskal–Wallis test; **e**, one-way ANOVA with Dunnett’s test; see Supplementary Information for exact P values).



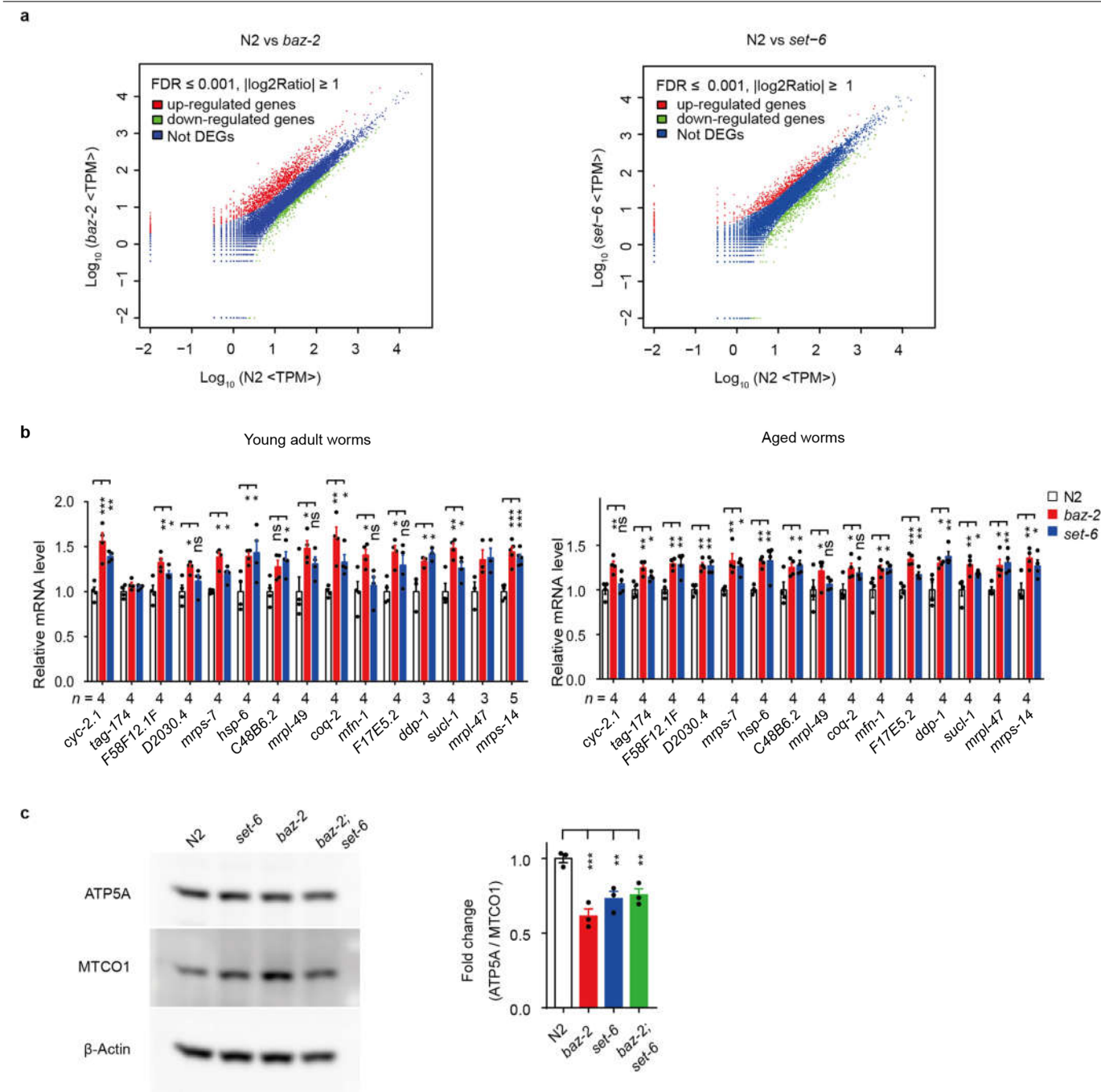
Extended Data Fig. 3 | Deletion of *baz-2* or *set-6* extends lifespan and enhances stress resistance via mechanisms related to dietary restriction and mitochondrial function. a–c, Percentage survival of N2, *baz-2*, *set-6*, and *baz-2;set-6* worms under oxidative (a), ultraviolet (b) and heat (c) stress. **d**, Lifespan curves of *daf-16*, *daf-16;baz-2*, and *daf-16;set-6* mutant worms (left) and their abilities to resist to oxidative stress (right). **e**, Lifespan curves of *eat-2*, *eat-2;baz-2*, and *eat-2;set-6* mutant worms (left) and their abilities to resist to oxidative stress (right). **f**, Lifespan curves of *clk-1*, *clk-1;baz-2*, and *clk-1;set-6*

mutant worms (left) and their abilities to resist to oxidative stress (right). In a–f, for oxidative-stress assays, data shown are means \pm s.e.m.; one-way ANOVA with Dunnett's test; numbers of independent experiments are shown beneath the bars. For heat-shock, ultraviolet-stress and lifespan assays, data represent the sum of animals in multiple experiments; two-sided log-rank test. The numbers of independent experiments and of tested hermaphrodites are indicated in parentheses. In all assays, * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ns, not significant (see Supplementary Information for exact P values).



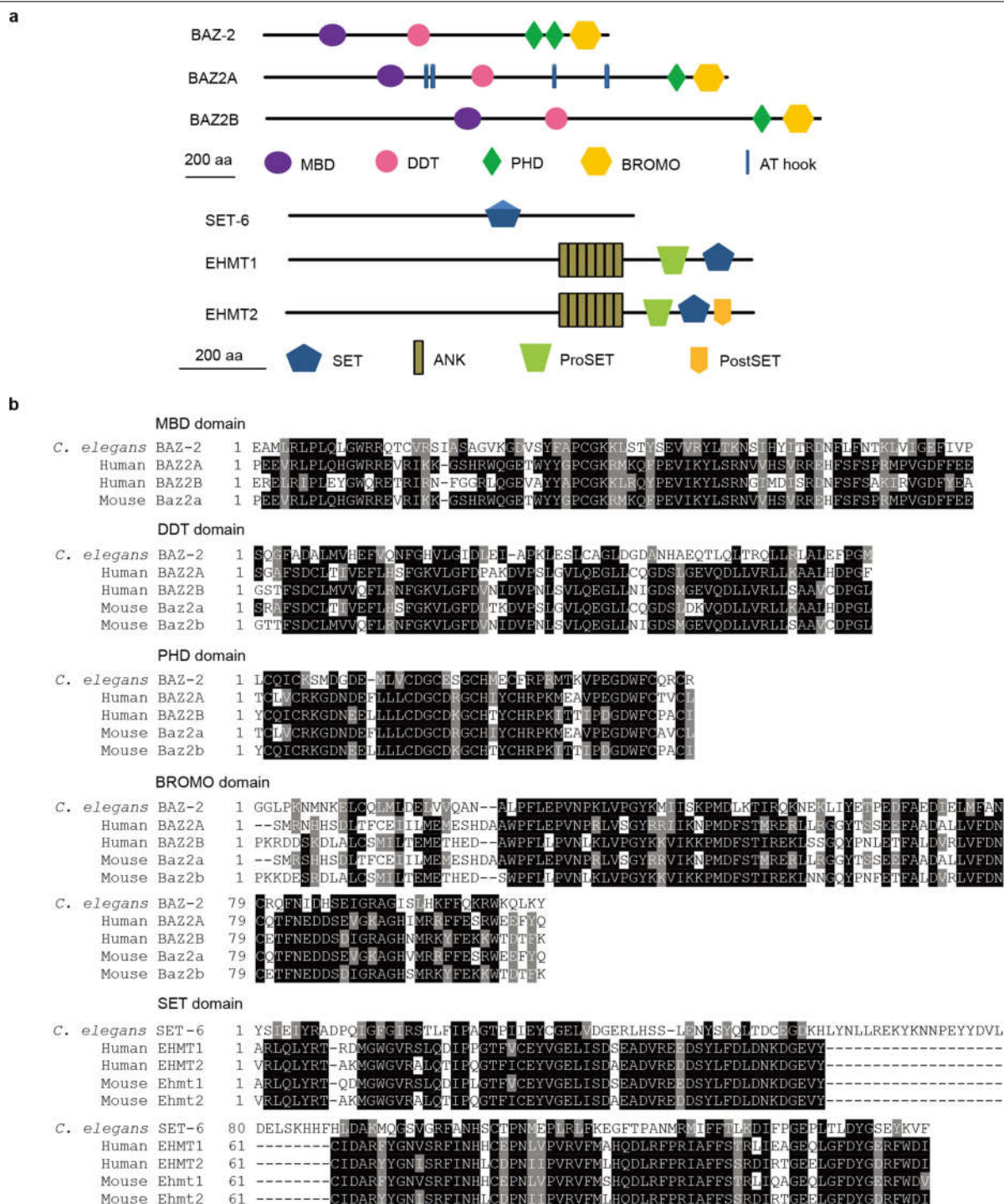
Extended Data Fig. 4 | Epigenetic regulators BAZ-2 and SET-6 localize at the promoter region of target genes. **a**, Co-immunoprecipitation of BAZ-2 and SET-6 using genome-edited *baz-2^{GFP::FLAG};set-6^{GFP::HA}* worms (left) or transgenic worms expressing BAZ-2::FLAG and SET-6::HA (right). Images are representative of four independent experiments. For gel source data, see Supplementary Fig. 3. **b**, Representative western blots (left) and quantitative analysis (right) of H3K9 methylation levels in N2, *baz-2*, *set-6* and *baz-2;set-6* worms. Normalized H3K9 methylation levels were calculated by normalizing the ratio of H3K9 methylation and histone 3 levels to that of N2 worms. For gel source data, see Supplementary Fig. 4. **c**, Peaks of BAZ-2- and SET-6-binding sites in the region -1,000 bp to +1000 bp around the transcription start site

(TSS). Only those peaks with a fold change of more than 2 are plotted. The y-axis indicates the average read coverage normalized to the number of uniquely mapped reads per million per genomic bin (bins = 1,000). **d**, Pie chart showing the distribution of overlapping BAZ-2- and SET-6-binding sites in genomic features. TTS, transcription termination site. **e**, **f**, ChIP-qPCR analysis of endogenous BAZ-2 (**e**) or SET-6 (**f**) enrichment at nuclear genes encoding mitochondrial proteins in genome-edited *baz-2^{GFP::FLAG}* and *set-6^{GFP::HA}* worms. ChIP-qPCR data from N2 worms were used as a control. In **b**, **e**, **f**, the numbers of independent experiments are shown beneath the bars; data are means \pm s.e.m.; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (**b**, one-way ANOVA with Dunnett's test; **e**, **f**, two-tailed *t*-test; see Supplementary Information for exact *P* values).



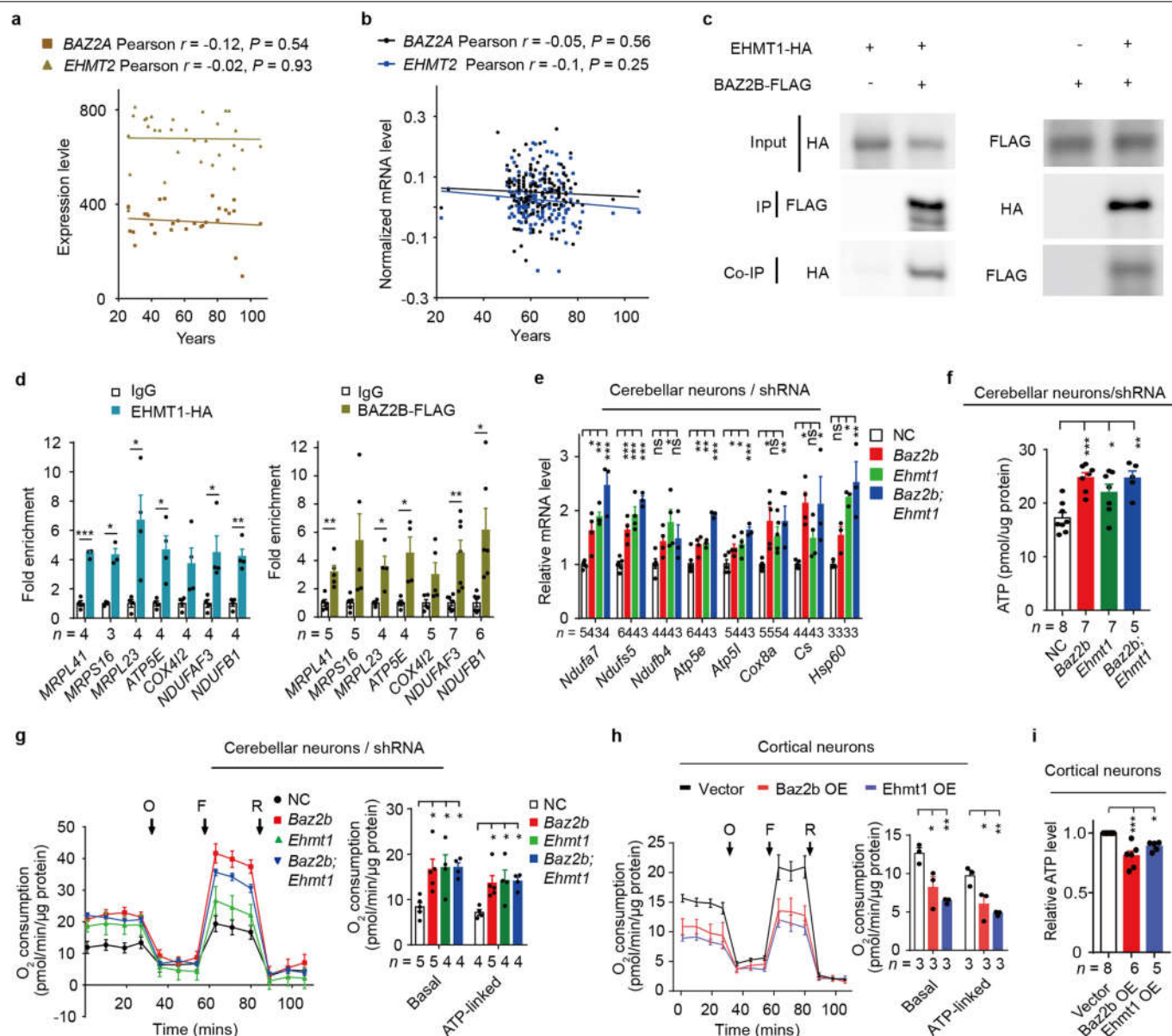
Extended Data Fig. 5 | BAZ-2 and SET-6 regulate the expression of nuclear genes encoding mitochondrial proteins. a, Scatter plots of mRNA-seq data for N2 versus *baz-2* mutant worms (left) and N2 versus *set-6* mutant worms (right). The x- and y-axes represent the log₁₀-transformed transcripts per million clean tags (TPM) expression values of N2 (x-axis), *baz-2* (left, y-axis) and *set-6* (right, y-axis) animals. Differentially expressed genes (DEGs) were defined through the parameters of false discovery rate (FDR) \leq 0.001 and $|\log_2\text{Ratio}| \geq 1$. **b**, RT-PCR analysis of changes in the expression of nuclear genes encoding

mitochondrial proteins in *baz-2* and *set-6* mutant worms at day 1 (left) or day 7 (right) of adulthood. The numbers of independent experiments are shown beneath the bars. **c**, Western blots (left) and quantitative analysis (right) of nDNA-encoded ATP5A and mtDNA-encoded MTCO1 proteins. $n=3$ independent experiments. For gel source data, see Supplementary Fig. 5. In **b**, **c**, data shown are means \pm s.e.m.; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; ns, not significant (one-way ANOVA with Dunnett's test; see Supplementary Information for exact P values).



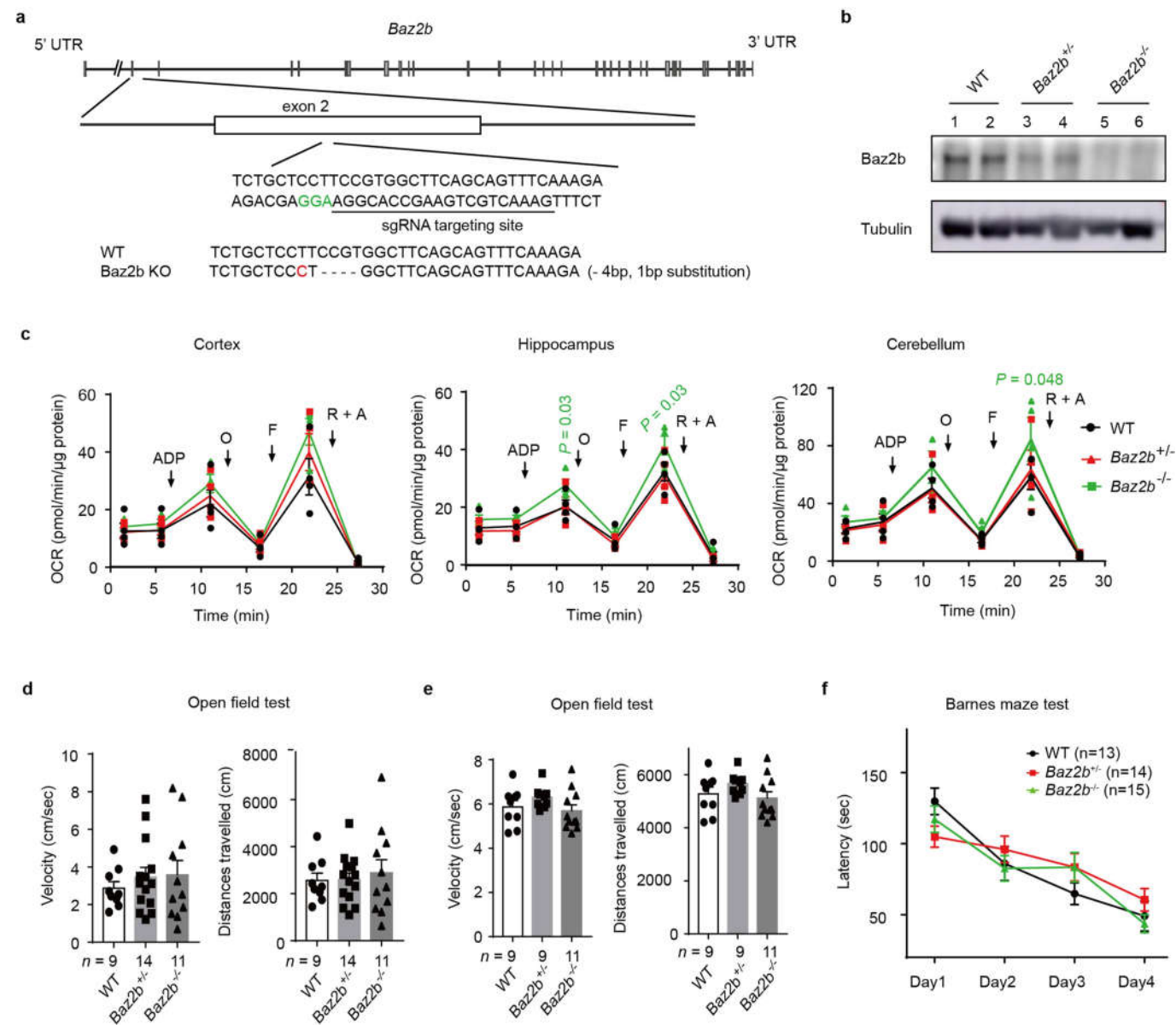
Extended Data Fig. 7 | Human homologues of BAZ-2 and SET-6. a, Diagram showing the similarity between *C. elegans* BAZ-2 and SET-6 and their human homologues. aa, amino acids; ANK, ankyrin repeats; BROMO, bromo domain; DDT, 'DNA-binding homeobox and different transcription factors' domain;

MBD, methyl-CpG binding domain; PHD, plant homeodomain; SET, Su(var.)3-9, Enhancer-of-zeste, Trithorax domain; **b**, Alignment of conserved domains in BAZ-2 or SET-6 with those in their mammalian homologues. Identical and conservative residues are highlighted in black and in grey, respectively.



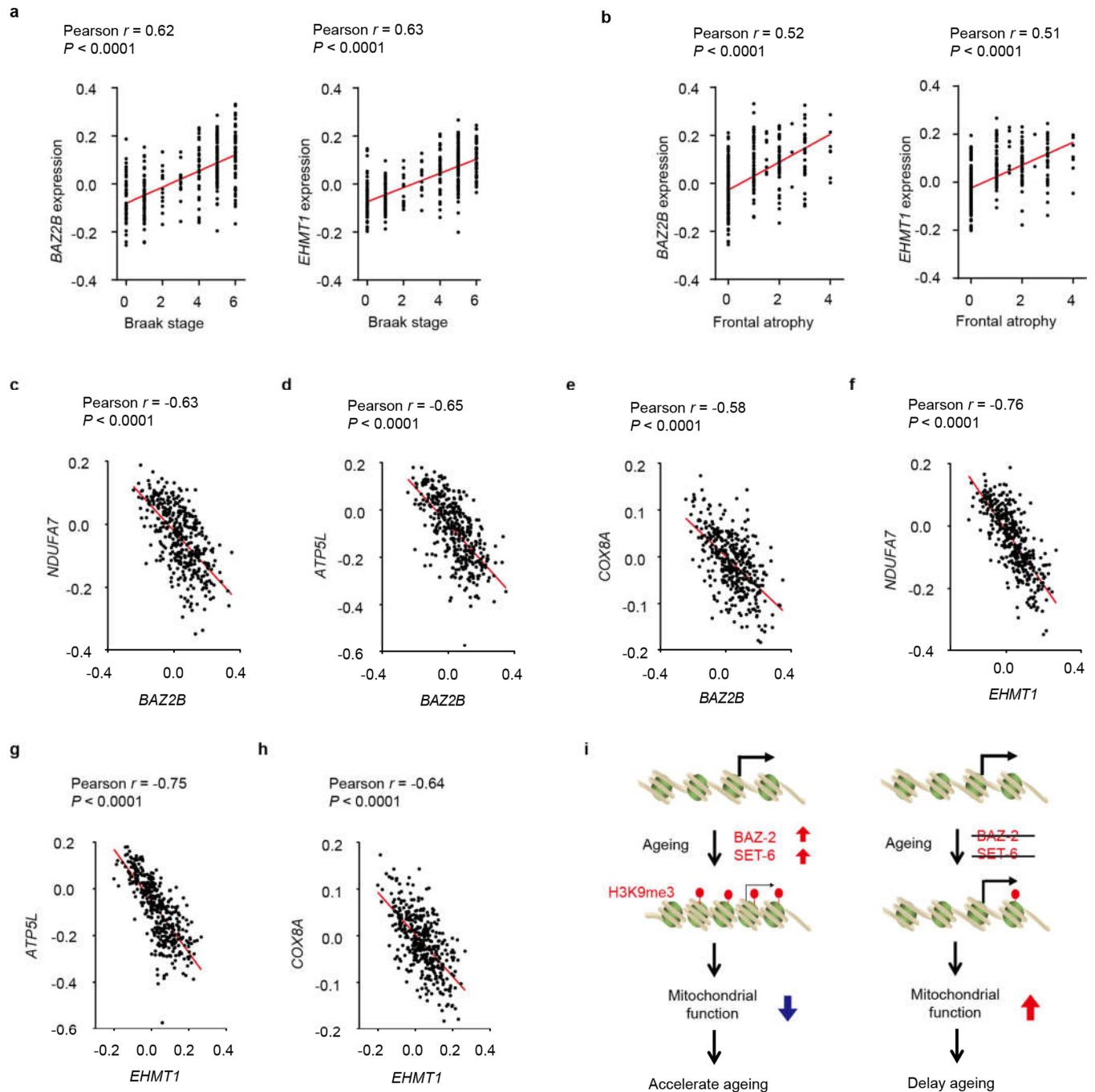
Extended Data Fig. 8 | Mammalian *Baz2b* and *Ehmt1* have a conserved role in repressing mitochondrial function. **a, b**, Transcription levels of *BAZ2A* and *EHMT2* in the prefrontal cortex of human brains at different ages. Expression values in **a** ($n = 30$ samples) are from the dataset GSE1572 (see ref. ⁸), and in **b** ($n = 145$ samples) are from brain samples without neurodegenerative disease in the dataset GSE44772 (see ref. ⁹). Pearson's r correlation coefficient was used for statistical testing. **c**, Co-immunoprecipitation of *BAZ2B* and *EHMT1* in HEK293T cells. Images are representative of three independent experiments. For gel source data, see Supplementary Fig. 6. **d**, ChIP-qPCR analysis of *BAZ2B* or *EHMT1* enrichment at nuclear genes encoding mitochondrial proteins in HEK293T cells. Immunoglobulin G (IgG) antibody was used as a control.

e–g, Effects of downregulation of mouse *Baz2b* or *Ehmt1* by short hairpin RNAs (shRNAs) on the transcription of mitochondria-related genes (**e**), ATP level (**f**) and oxygen-consumption rate (**g**) in primary mouse cerebellar neurons. F, carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP); NC, negative control; O, oligomycin; R, rotenone. **h, i**, Oxygen-consumption rate (**h**) and ATP content (**i**) in primary mouse cortical neurons overexpressing (OE) mouse *Baz2b* or *Ehmt1*. Graphical OCR data are representative of three independent experiments. In **d–i**, sample numbers are shown beneath the bars. Data shown are means \pm s.e.m.; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (**d**, two-tailed t -test, **e–h**, one-way ANOVA with Dunnett's test; **i**, Kruskal–Wallis test; see Supplementary Information for exact P values).



Extended Data Fig. 9 | Effects of deleting *Baz2b* on mitochondrial function and mouse behaviour. **a**, Diagram showing the generation of *Baz2b*^{-/-} (KO) mice. UTR, untranslated repeat; WT, wild-type. Protospacer-adjacent motif (PAM) sequences are highlighted in green; the substitution site is highlighted in red. **b**, Western blot analysis of *Baz2b* protein level in samples from WT, *Baz2b*^{+/+} and *Baz2b*^{-/-} mice. Images are representative of three independent experiments. For gel source data, see Supplementary Fig. 7. **c**, Oxygen-consumption rates (OCRs) of mitochondria isolated from 12-month-old WT,

Baz2b^{+/+} and *Baz2b*^{-/-} male mice. A, antimycin; F, carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP); O, oligomycin; R, rotenone. *n* = 4 mice per group. *P* values were determined by one-way ANOVA with Dunnett's test. **d**, **e**, Spontaneous locomotion of old (**d**) and young (**e**) mice in an open field test. The numbers of tested mice are shown beneath the bars. **f**, Escape latency in Barnes maze trials during training days for young WT, *Baz2b*^{+/+} and *Baz2b*^{-/-} mice. Numbers of tested mice are indicated in parentheses. In all assays, data shown are means ± s.e.m.



Extended Data Fig. 10 | Expression level of *BAZ2B* and *EHMT1* correlates positively with the progression of Alzheimer's disease in humans.

a, Expression level of *BAZ2B* (left) and *EHMT1* (right) at different Braak stages in the prefrontal cortex of brains with Alzheimer's disease. **b**, Expression level of *BAZ2B* (left) and *EHMT1* (right) at different stages of frontal atrophy in the prefrontal cortex of brains with Alzheimer's disease. **c–e**, Correlations between the expression levels of *BAZ2B* (x-axes) and selected nuclear genes (encoding mitochondrial proteins; y-axes) in the prefrontal cortex of brains with

Alzheimer's disease. **f–h**, Correlations between the expression levels of *EHMT1* (x-axes) and selected nuclear genes (encoding mitochondrial proteins; y-axes) in the prefrontal cortex of brains with Alzheimer's disease. In **a–h**, expression values ($n = 390$ samples) of examined genes are from the dataset GSE44772, using Pearson's r correlation coefficient for statistical testing. **i**, Proposed working model for the epigenetic regulation of mitochondrial function and healthy ageing.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	All confocal images were taken by the Nikon A1 with the NIS elements software.
Data analysis	<p>ChIP-seq analysis: cutadapt (version 1.15), Bowtie 2 (version 2.3.3.1), MACS (version 2.1.0.20150731), HOMER (version 4.9, 2-20-2017), samtools (version 1.5), metaseq (version 0.5.5.4), bedtools (version 2.26.0), deeptools (version 2.5.4).</p> <p>GO enrichment analysis: DAVID website (https://david.ncifcrf.gov/, version 6.8), seaborn (version 0.8.1).</p> <p>RNA-seq analysis: cutadapt (version 1.10), HISAT2 (version 2.0.5), HTSeq (version 0.9.1), DESeq2 (version 1.18.1).</p> <p>GraphPad Prism 7.0 was utilized for statistical analysis and generating graphs.</p> <p>Image J was used for image quantitation.</p> <p>Gene network analysis: GeneMANIA.</p> <p>Novel location recognition test was analysed using the software package developed by Nirao M. Shah's lab.</p> <p>Open field test was analysed by EthoVision XT (version 11.5).</p> <p>Custom codes are all available in GitHub (https://github.com/SHYKON-YIN/nature)</p>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw sequenced data generated in this study were deposited in the NCBI Sequence Read Archive under accession number PRJNA554977. Figure 3b-d, Extended

Data Figure 4c-d, Extended Data Figure 5a, and Extended data Figure 6a-d are associated with the raw data. All other data supporting the findings of this study are available within the paper and its supplementary information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes for in vitro experiments. In our experiments, sample sizes were estimated based on previous literatures and were chosen in order to be able to perform statistical analyses, as is standard in the field.
Data exclusions	In the lifespan and RNAi assays, worms that crawled off the plate were not included in the data. For the male mating assay, mating plates with males or hermaphrodites that crawled out the plate during the assay were not included in the analysis. These criteria were established before starting the experiments. For the novel location test, mice that do not explore more than 3 sec total for both objects during training or testing are excluded from analysis. Animals that have discrimination indexes >0.2 or <-0.2 at training are considered to have a significant location/object bias during training and are also excluded from analysis.
Replication	All attempts at replication were successful
Randomization	Worms and mice were chosen unbiasedly for experimental analysis to ensure randomization
Blinding	The investigators were blinded to the genotype or RNAi treatments during the lifespan and behavioural assays

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

- 1) Mouse monoclonal anti-Flag-Tag antibody (Same as Sigma's Anti-FLAG?): Clone/3B9; Supplier/Abmart; Cat.No./ M20008; Lot.No./283658; Dilution: 1/1000 for western blot.
- (2) Rat monoclonal anti-HA antibody: Clone/3F10; Supplier/Roche; Cat.No./11867431001; Lot.No./15645900 and 12177700; Dilution: 1/3000 for western blot.
- (3) Rabbit polyclonal antibody for detection of Histone H3 monomethylated on lysine 9 (H3K9me1): Supplier/abcam; Cat.No./ ab9045; Lot.No./GR3198013-1, Dilution: 1/1000 for western blot. Use 5 ug for 10 µg of chromatin in ChIP experiment.
- (4) Mouse monoclonal antibody to Histone H3 dimethylation on lysine 9 (H3K9me2): Supplier/Abcam; Cat.No./ab1220; Lot.No./ GR325223-4; Use 5 ug for 10 µg of chromatin in ChIP experiment.
- (5) Rabbit polyclonal antibody to Histone H3 dimethylation on lysine 9 (H3K9me2): Supplier/Abcam; Cat.No./ ab115159; Dilution: 1/1000 for western blot.
- (6) Rabbit polyclonal antibody to Histone H3 (tri methyl K9) antibody: Supplier/Abcam; Cat.No./ab8898; Lot.No./GR3217826-1 and GR148830-2; Dilution: 1/1000 for western blot. Use 5 ug for 10 µg of chromatin in ChIP experiment.
- (7) Rabbit polyclonal antibody to Histone H3: Supplier/Abcam; Cat.No./ab1791; Lot.No./GR94293-1; Dilution: 1/10000 for western blot.
- (8) Mouse monoclonal anti-FLAG M2 antibody: Clone/ M2; Supplier/SIGMA; Cat.No./F3165; Lot.No./SLBQ7119V; Dilution: 1/1000 for western blot; Use 2 µg for 1 µg of chromatin in ChIP experiment.
- (9) Mouse monoclonal anti GLP/EHMT 1 Antibody: Clone/B0422; Supplier/ R&D Systems; Cat.No./PP-B0422-00; Lot.No./A-2; Dilution: 1/1000 for western blot; Use 2 µg for 1 µg of chromatin in ChIP experiment.

(10) Monoclonal Anti- α -Tubulin antibody produced in mouse: clone/B-5-1-2; Supplier/SIGMA; Cat.No./T6074; Dilution: 1/5000 for western blot.
 (11) Rabbit polyclonal antibody against Baz2b were generated by collaboration with Abcam; not commercially available; Dilution: 1/1500 for western blot.
 (12) ATP5A: Anti-ATP5A antibody [15H4C4] - Mitochondrial Marker: Supplier/Abcam; Cat.No./ ab14748 ; Lot.No./GR209582-8; Dilution: 1/1000 for western blot.
 (13) MTCO1: Anti-MTCO1 antibody [1D6E1A8]; Supplier/Abcam; Cat.No./ ab14705 ; Lot.No./GR233531-1; Dilution: 1/1000 for western blot.
 (14) Rabbit polyclonal antibody to β -Actin: Supplier/Abmart; Cat.No./P30002M; Lot.No./294357; Dilution: 1/1000 for western blot.
 (15) IRDye 800CW Goat anti-Rabbit IgG Secondary Antibody: Supplier/ LI-COR Biosciences; Cat.No./ P/N 925-32211; Dilution: 1/5000 for western blot.
 (16) IRDye 680RD Goat anti-Rabbit IgG Secondary Antibody: Supplier/ LI-COR Biosciences; Cat.No./ P/N 925-68071; Dilution: 1/10000 for western blot.

Validation

Rabbit polyclonal against Baz2b antibody was validated using our Baz2b knock-out mice, the result was shown in this manuscript. All the other antibodies used in this study were validated by the suppliers, the information was available on their website.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293T (Catalogue number SCSP-502) cell line was ordered from the cell bank of the Chinese Academy of Sciences.
Authentication	The cell line has been validated using the short tandem repeat (STR) profiling method by the cell bank of Chinese Academy of Sciences.
Mycoplasma contamination	Not tested
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	C. elegans strains: The wild-type Bristol N2, set-6(ok2195), daf-16(mu86), eat-2(ad1116), clk-1(qm30), and TU3401 (sid-1(pk3321);uls69) strains were obtained from the Caenorhabditis Genetics Center. baz-2(tm0235) was obtained from National Bioresource Project, Japan. set-6;baz-2 double mutant worms were generated by crossing set-6(ok2195) with baz-2(tm0235). The TU3401;Pbas-1::bas-1::gfp strain was generated by crossing TU3401 with Pbas-1::bas-1::gfp transgenic worms (SQ0017). The baz-2;Pbas-1::bas-1::gfp and set-6;Pbas-1::bas-1::gfp strains were generated by crossing baz-2(tm0235) and set-6(ok2195) with Pbas-1::bas-1::gfp worms, respectively. The baz-2;daf-16 and set-6;daf-16 mutant worms were generated by crossing daf-16(mu86) with baz-2(tm0235) and set-6(ok2195), respectively. The baz-2;eat-2 and set-6;eat-2 mutant worms were generated by crossing eat-2(ad1116) with baz-2(tm0235) and set-6(ok2195), respectively. The locus of the baz-2 is adjacent to that of clk-1, so the baz-2;clk-1 strain was generated by deleting baz-2 (yfh0100 allele) in the genomic background of clk-1(qm30) mutant worms using the CRISPR-Cas9 system. The set-6;clk-1 worms were generated by crossing set-6(ok2195) with clk-1(qm30). The Phsp-6::gfp transgenic strain SJ4100 was crossed with set-6(ok2195), baz-2(tm0235), and set-6;baz-2 worms to express Phsp-6::GFP in these mutant worms. C57BL/6J at embryonic day 14.5 and post-natal day 6 were obtained from Shanghai Laboratory Animal Center, Chinese Academy of Sciences. The null mutation in Baz2b of C57BL/6J mice were generated by Suzhou Non-human Primate Facility, Institute of Neuroscience, Chinese Academy of Sciences. Both male and female mice were used in this study. Breeding, housing, and experimental procedures were performed following the procedure approved by the Institutional Animal Care and Use Committee at the Institute of Neuroscience, Chinese Academy of Sciences.
Wild animals	This study did not involve wild animals.
Field-collected samples	Not involve
Ethics oversight	All procedures for performing mouse behavioural test were approved by the Animal Care and Use Committee of the Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

The raw sequenced data generated in this study were deposited in the NCBI Sequence Read Archive under accession number PRJNA554977.

Files in database submission

N2-a2-H3K9me1_ChIP
 N2-a2-H3K9me1_Input
 N2-a2-H3K9me2_ChIP
 N2-a2-H3K9me2_Input
 N2-a2-H3K9me3_ChIP
 N2-a2-H3K9me3_Input
 BAZ2-a2-H3K9me1_ChIP
 BAZ2-a2-H3K9me1_Input
 BAZ2-a2-H3K9me2_ChIP
 BAZ2-a2-H3K9me2_Input
 BAZ2-a2-H3K9me3_ChIP
 BAZ2-a2-H3K9me3_Input
 SET6-a2-H3K9me1_ChIP
 SET6-a2-H3K9me1_Input
 SET6-a2-H3K9me2_ChIP
 SET6-a2-H3K9me2_Input
 SET6-a2-H3K9me3_ChIP
 SET6-a2-H3K9me3_Input
 N2-a7-H3K9me1_ChIP
 N2-a7-H3K9me1_Input
 N2-a7-H3K9me2_ChIP
 N2-a7-H3K9me2_Input
 N2-a7-H3K9me3_ChIP
 N2-a7-H3K9me3_Input
 BAZ2-a7-H3K9me1_ChIP
 BAZ2-a7-H3K9me1_Input
 BAZ2-a7-H3K9me2_ChIP
 BAZ2-a7-H3K9me2_Input
 BAZ2-a7-H3K9me3_ChIP
 BAZ2-a7-H3K9me3_Input
 SET6-a7-H3K9me1_ChIP
 SET6-a7-H3K9me1_Input
 SET6-a7-H3K9me2_ChIP
 SET6-a7-H3K9me2_Input
 SET6-a7-H3K9me3_ChIP
 SET6-a7-H3K9me3_Input
 N2-a2-H3K9me3-replicate_ChIP
 N2-a2-H3K9me3-replicate_Input
 N2-a7-H3K9me3-replicate_ChIP
 N2-a7-H3K9me3-replicate_Input
 BAZ2-a2-H3K9me3-replicate_ChIP
 BAZ2-a2-H3K9me3-replicate_Input
 BAZ2-a7-H3K9me3-replicate_ChIP
 BAZ2-a7-H3K9me3-replicate_Input
 SET6-a2-H3K9me3-replicate_ChIP
 SET6-a2-H3K9me3-replicate_Input
 SET6-a7-H3K9me3-replicate_ChIP
 SET6-a7-H3K9me3-replicate_Input

 SET6-GFP_ChIP
 SET6-GFP_Input
 BAZ2-GFP_ChIP
 BAZ2-GFP_Input

Genome browser session
 (e.g. [UCSC](#))

The raw sequenced data generated in this study were deposited in the NCBI Sequence Read Archive under accession number PRJNA554977.

Methodology

Replicates

There is one ChIP-seq experiment of BAZ-2/SET-6 binding, H3K9m1, H3K9me2 and two replicates for ChIP-seq experiment of H3K9me3. Three batches of worms were collected for each ChIP-seq measurement.

Sequencing depth

We performed the Illumina single-end 50bp sequencing. The sequence depth of each sample was described as following:
 BAZ-2_GFP-ChIP Total reads: 27896954 Uniquely mapped reads: 20776365
 BAZ-2_GFP-Input Total reads: 41170127 Uniquely mapped reads: 32719966
 SET-6_GFP-ChIP Total reads: 26454834 Uniquely mapped reads: 18360852
 SET-6_GFP-Input Total reads: 32004369 Uniquely mapped reads: 24228355
 We performed the Illumina pair-end 150bp sequencing. The sequence depth of each sample was described as following:
 N2-a2-H3K9me1_ChIP Total reads: 34729875 Uniquely mapped reads: 27623314
 N2-a2-H3K9me1_Input Total reads: 39345751 Uniquely mapped reads: 30387782
 N2-a2-H3K9me2_ChIP Total reads: 28468453 Uniquely mapped reads: 19968945

N2-a2-H3K9me2_Input Total reads: 26552809 Uniquely mapped reads: 21519737
 N2-a2-H3K9me3_ChIP Total reads: 28765524 Uniquely mapped reads: 21960227
 N2-a2-H3K9me3_Input Total reads: 26552809 Uniquely mapped reads: 21519737
 BAZ2-a2-H3K9me1_ChIP Total reads: 36799862 Uniquely mapped reads: 29242685
 BAZ2-a2-H3K9me1_Input Total reads: 28694706 Uniquely mapped reads: 22470302
 BAZ2-a2-H3K9me2_ChIP Total reads: 27119242 Uniquely mapped reads: 19672636
 BAZ2-a2-H3K9me2_Input Total reads: 29085145 Uniquely mapped reads: 22446541
 BAZ2-a2-H3K9me3_ChIP Total reads: 31181664 Uniquely mapped reads: 24045304
 BAZ2-a2-H3K9me3_Input Total reads: 29085145 Uniquely mapped reads: 22446541
 SET6-a2-H3K9me1_ChIP Total reads: 39567207 Uniquely mapped reads: 31080460
 SET6-a2-H3K9me1_Input Total reads: 32115006 Uniquely mapped reads: 25974182
 SET6-a2-H3K9me2_ChIP Total reads: 27634238 Uniquely mapped reads: 19793956
 SET6-a2-H3K9me2_Input Total reads: 29907795 Uniquely mapped reads: 23793106
 SET6-a2-H3K9me3_ChIP Total reads: 26680288 Uniquely mapped reads: 20452910
 SET6-a2-H3K9me3_Input Total reads: 29907795 Uniquely mapped reads: 23793106
 N2-a7-H3K9me1_ChIP Total reads: 15889684 Uniquely mapped reads: 12872030
 N2-a7-H3K9me1_Input Total reads: 34280879 Uniquely mapped reads: 27097216
 N2-a7-H3K9me2_ChIP Total reads: 33272702 Uniquely mapped reads: 23613681
 N2-a7-H3K9me2_Input Total reads: 39690761 Uniquely mapped reads: 32441690
 N2-a7-H3K9me3_ChIP Total reads: 42692956 Uniquely mapped reads: 31858518
 N2-a7-H3K9me3_Input Total reads: 39690761 Uniquely mapped reads: 32441690
 BAZ2-a7-H3K9me1_ChIP Total reads: 36216152 Uniquely mapped reads: 29353294
 BAZ2-a7-H3K9me1_Input Total reads: 36518674 Uniquely mapped reads: 27660179
 BAZ2-a7-H3K9me2_ChIP Total reads: 35811329 Uniquely mapped reads: 24767427
 BAZ2-a7-H3K9me2_Input Total reads: 39064259 Uniquely mapped reads: 31913061
 BAZ2-a7-H3K9me3_ChIP Total reads: 31221555 Uniquely mapped reads: 22026684
 BAZ2-a7-H3K9me3_Input Total reads: 39064259 Uniquely mapped reads: 31913061
 SET6-a7-H3K9me1_ChIP Total reads: 16264960 Uniquely mapped reads: 13156364
 SET6-a7-H3K9me1_Input Total reads: 36626677 Uniquely mapped reads: 29121215
 SET6-a7-H3K9me2_ChIP Total reads: 33701820 Uniquely mapped reads: 23610207
 SET6-a7-H3K9me2_Input Total reads: 43091463 Uniquely mapped reads: 35286028
 SET6-a7-H3K9me3_ChIP Total reads: 38185317 Uniquely mapped reads: 28398527
 SET6-a7-H3K9me3_Input Total reads: 43091463 Uniquely mapped reads: 35286028
 N2-a2-H3K9me3-replicate_ChIP Total reads: 62995763 Uniquely mapped reads: 60135755
 N2-a2-H3K9me3-replicate_Input Total reads: 50887869 Uniquely mapped reads: 48979574
 N2-a7-H3K9me3-replicate_ChIP Total reads: 41419670 Uniquely mapped reads: 40135660
 N2-a7-H3K9me3-replicate_Input Total reads: 48541033 Uniquely mapped reads: 34973814
 BAZ2-a2-H3K9me3-replicate_ChIP Total reads: 53384244 Uniquely mapped reads: 51462411
 BAZ2-a2-H3K9me3-replicate_Input Total reads: 37493544 Uniquely mapped reads: 36256257
 BAZ2-a7-H3K9me3-replicate_ChIP Total reads: 64916141 Uniquely mapped reads: 62942690
 BAZ2-a7-H3K9me3-replicate_Input Total reads: 35799538 Uniquely mapped reads: 24687361
 SET6-a2-H3K9me3-replicate_ChIP Total reads: 53978506 Uniquely mapped reads: 52229602
 SET6-a2-H3K9me3-replicate_Input Total reads: 59507865 Uniquely mapped reads: 54687728
 SET6-a7-H3K9me3-replicate_ChIP Total reads: 40578210 Uniquely mapped reads: 39186377
 SET6-a7-H3K9me3-replicate_Input Total reads: 38282802 Uniquely mapped reads: 35132127

Antibodies	GFP-trap agarose beads (Chromotec ACT-CM-GFA0250); H3K9me1 (Abcam ab9045), H3K9me2 (Abcam ab1220) and H3K9me3 (Abcam ab8898)
Peak calling parameters	<p>All reads including ChIP and Input were aligned to the reference worm genome(WormBase, WS256) using the Bowtie 2 aligner (version 2.3.3.1) with the following parameters: bowtie2 -p 6 -N 1 -x \$index -U \$read -S \$sam >> \$log 2>&1</p> <p>For peak calling, we used the MACS software (version 2.1.0.20150731) with the following parameters: macs2 callpeak -t \$ChIP -c \$Input -n \$name -g ce --outdir \$out -f BAM -p 0.05 -B;</p> <p>For bedgraph files, we used the following parameters: bamCoverage -b \$file -p 30 -e 300 --binSize 10 --centerReads --normalizeUsing RPKM -o \${file%.*}.rpkm.bg</p>
Data quality	<p>Sequencing data quality was checked with the fastqc software. Quality of the alignment was checked with samtools flagstat. We optimized the fold enrichment cutoff for peaks selecting and finally identified 5714 significantly BAZ-2 occupied loci upon filtering high quality peaks based on at least 2-fold enrichment and less than 5% FDR. Using the same criteria, we identified 5198 SET-6 peaks. For H3K9me1/2/3 histone modification ChIP-seq data, we first mapped to reference worm genome(WormBase, WS256) and then filtered out PCR duplicates and low quality reads. Finally, we generated bedgraph files using 10bp bins.</p>
Software	<p>Cutadapt (version 1.15) was used to trim adapter sequence. Bowtie2 aligner (version 2.3.3.1) was used to align reads to genome. SAMtools (version 1.5) was used to filter the SAM and BAM file to get uniquely mapped reads. HOMER (version 4.9, 2-20-2017) annotatePeaks.pl function was used to annotate peaks file. The intersect function of bedtools (version 2.26.0) was used to obtain the overlapping peaks of BAZ-2 and SET-6. The bamCoverage function of deeptools (version 2.5.4) The metaseq (version 0.5.5.4) framework with the RPM normalization methods was used to plot the reads average coverage profile of all ChIP-seq datasets.</p>


Global chemical effects of the microbiome include new bile-acid conjugations

<https://doi.org/10.1038/s41586-020-2047-9>

Received: 6 July 2018

Accepted: 3 January 2020

Published online: 26 February 2020

 Check for updates

Robert A. Quinn^{1,2}, Alexey V. Melnik¹, Alison Vrbancac³, Ting Fu⁴, Kathryn A. Patras³, Mitchell P. Christy¹, Zsolt Bodai⁵, Pedro Belda-Ferre³, Anupriya Tripathi^{1,3}, Lawton K. Chung³, Michael Downes⁴, Ryan D. Welch⁴, Melissa Quinn⁶, Greg Humphrey³, Morgan Panitchpakdi¹, Kelly C. Weldon^{1,9}, Alexander Aksenov¹, Ricardo da Silva¹, Julian Avila-Pacheco⁷, Clary Clish⁷, Sena Bae^{8,9}, Himel Mallick^{7,8}, Eric A. Franzosa^{7,8}, Jason Lloyd-Price^{7,8}, Robert Bussell¹⁰, Taren Thron¹¹, Andrew T. Nelson¹, Mingxun Wang¹, Eric Leszczynski⁶, Fernando Vargas¹, Julia M. Gauglitz¹, Michael J. Meenan¹, Emily Gentry¹, Timothy D. Arthur^{3,7}, Alexis C. Komor⁵, Orit Poulsen³, Brigid S. Boland¹², John T. Chang¹², William J. Sandborn¹², Meerana Lim³, Neha Garg^{13,14}, Julie C. Lumeng¹⁵, Ramnik J. Xavier⁷, Barbara I. Kazmierczak¹⁶, Ruchi Jain¹⁶, Marie Egan¹⁷, Kyung E. Rhee³, David Ferguson⁶, Manuela Raffatellu³, Hera Vlamakis⁷, Gabriel G. Haddad³, Dionicio Siegel¹, Curtis Huttenhower^{7,8}, Sarkis K. Mazmanian¹¹, Ronald M. Evans^{4,18}, Victor Nizet^{1,3,19}, Rob Knight^{3,19,20,21} & Pieter C. Dorrestein^{1,3,19} ✉

A mosaic of cross-phylum chemical interactions occurs between all metazoans and their microbiomes. A number of molecular families that are known to be produced by the microbiome have a marked effect on the balance between health and disease^{1–9}. Considering the diversity of the human microbiome (which numbers over 40,000 operational taxonomic units¹⁰), the effect of the microbiome on the chemistry of an entire animal remains underexplored. Here we use mass spectrometry informatics and data visualization approaches^{11–13} to provide an assessment of the effects of the microbiome on the chemistry of an entire mammal by comparing metabolomics data from germ-free and specific-pathogen-free mice. We found that the microbiota affects the chemistry of all organs. This included the amino acid conjugations of host bile acids that were used to produce phenylalanocholic acid, tyroscholeic acid and leucocholic acid, which have not previously been characterized despite extensive research on bile-acid chemistry¹⁴. These bile-acid conjugates were also found in humans, and were enriched in patients with inflammatory bowel disease or cystic fibrosis. These compounds agonized the farnesoid X receptor in vitro, and mice gavaged with the compounds showed reduced expression of bile-acid synthesis genes in vivo. Further studies are required to confirm whether these compounds have a physiological role in the host, and whether they contribute to gut diseases that are associated with microbiome dysbiosis.

In total, we analysed 768 samples from 96 sample sites of 29 different organs from 4 germ-free and 4 colonized mice by liquid chromatography–tandem mass spectrometry (LC–MS/MS) and 16S rRNA gene sequencing (Supplementary Table 1). Mapping the first principal coordinate position of each sample from specific-pathogen-free (SPF) mice onto a three-dimensional (3D) mouse model¹³ enabled us to visualize the similarity of the microbiome and metabolome through all organs

and organ systems (Fig. 1a, b; the 3D model is available as Supplementary Data). Different sections through the gastrointestinal tract had unique microbiome and metabolome profiles. There was a distinct difference between the similarity of the two data types in mouse faecal samples. The metabolome differed between faecal samples and the distal gastrointestinal tract, whereas the microbiome was more similar between faeces and colon or caecum samples.

¹Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA, USA. ²Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, USA. ³Department of Pediatrics, University of California San Diego, San Diego, CA, USA. ⁴Gene Expression Laboratory, Salk Institute for Biological Studies, San Diego, CA, USA. ⁵Department of Chemistry and Biochemistry, University of California San Diego, San Diego, CA, USA. ⁶Department of Kinesiology, Michigan State University, East Lansing, MI, USA. ⁷Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁸Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA. ⁹Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA, USA. ¹⁰Department of Radiology, University of California San Diego, San Diego, CA, USA. ¹¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. ¹²Division of Gastroenterology, Department of Medicine, University of California San Diego, San Diego, CA, USA. ¹³School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA. ¹⁴Emory-Children's Cystic Fibrosis Center, Atlanta, GA, USA. ¹⁵Department of Pediatrics, University of Michigan, Ann Arbor, MI, USA. ¹⁶Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. ¹⁷Department of Pediatrics, Yale School of Medicine, New Haven, CT, USA. ¹⁸Howard Hughes Medical Institute, The Salk Institute for Biological Studies, San Diego, CA, USA. ¹⁹UCSD Center for Microbiome Innovation, University of California San Diego, San Diego, CA, USA. ²⁰Department of Computer Science and Engineering, University of California San Diego, San Diego, CA, USA.

²¹Department of Engineering, University of California San Diego, San Diego, CA, USA. ✉e-mail: pdorrestein@ucsd.edu

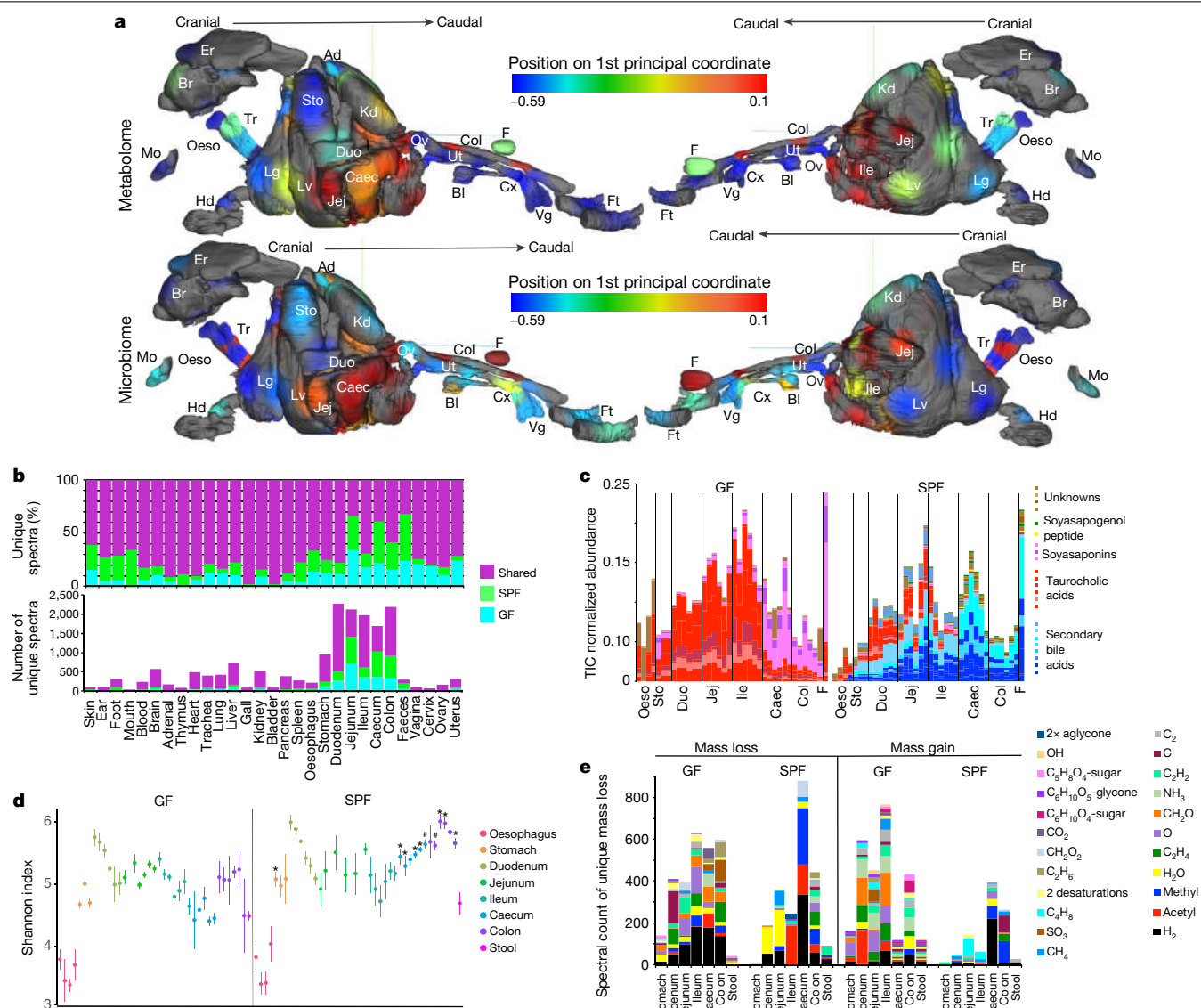


Fig. 1 | Global effect of the microbiome on the chemistry of an entire mammal. **a**, Three-dimensional model of mouse organs mapped with the mean first principal coordinate (Extended Data Fig. 1) as a heat map (according to the colour scale), from the germ-free and SPF mice ($n = 4$ mice each). Ad, adrenal gland; bl, bladder; br, brain; caec, caecum; col, colon; cx, cervix; duo, duodenum; er, ear; f, faeces; ft, feet; hd, hand; jej, jejunum; kd, kidney; lg, lung; lv, liver; mo, mouth; oes, oesophagus; ov, ovary; sto, stomach; tr, trachea; ut, uterus; vg, vagina. **b**, Mean percentage and total number of unique spectra in each organ sampled from the two mouse groups. **c**, Relative abundance (normalized to total ion current (TIC)) of the 30 most differential metabolites between the guts of germ-free and SPF mice. The metabolites are coloured as secondary bile acids (blue), primary bile acids (red), soyasaponins (pink),

Molecular networking of mouse data

To characterize the chemical effect of the microbiome, we subjected the mass spectrometry data to molecular networking¹². The algorithm identified 7,913 spectra, of which $14.7 \pm 2.2\%$ were observed in colonized mice and $10.0 \pm 0.7\%$ were exclusive to germ-free mice (Fig. 1c, Extended Data Fig. 1). Although the overall profiles showed that the strongest differences between germ-free and SPF mice were in the gastrointestinal tract, molecular networking identified unique chemical signatures from the microbiome in all organs—ranging from 2% in the bladder to 44% in stools (Fig. 1b). The metabolome of the caecum, the main site

of peptides (yellow) and unknown (brown). Annotations are based on spectral matching or molecular network propagation (level two or three¹⁶). Stereochemistry of the annotated molecules cannot be discerned using these methods. **d**, Mean and 95% confidence interval of the Shannon–Weiner diversity of the metabolomic data in each sample from the gastrointestinal tracts of germ-free and SPF mice. Statistical significance between metabolome diversity in the same sample location between germ-free and SPF mice was tested with a two-sided Mann–Whitney U -test, $n = 4$. * $P = 0.028$, * $P = 0.057$. **e**, Results of meta-mass-shift chemical profiling¹⁷ showing the spectral counts of known mass differences between unique nodes in either germ-free or SPF mice. Each mass difference corresponds to the node-to-node gain or loss of a particular chemical group.

of microbial fermentation of food, was most-markedly affected by the microbiota. Spectral library searching enabled the annotation of 8.9% of nodes in the molecular network^{11,15} (level two or three, according to previously published standards¹⁶). Many of the changes attributed to the microbiome were location-specific, resulting from the metabolism of plant natural products from food and bile acids (Fig. 1c, Extended Data Figs. 2–4, Supplementary Data).

In the upper gastrointestinal tract, the Shannon diversity of the metabolomes of germ-free mice mirrored those of SPF mice; in both sets of mice, diversity was low in the oesophagus and higher in the stomach and duodenum. Upon transition to the caecum, however,

the diversity of the two groups of mice began to separate (Fig. 1d). The molecular diversity in the caecum and colon of colonized mice was higher than that of germ-free mice, but this was not the case in the stool samples (Fig. 1d). In the duodenum (the location at which the gallbladder adds bile to the intestine), there was a contrast in microbiome and metabolome diversity: a high metabolome diversity corresponded to a low microbial diversity (Fig. 1d, Extended Data Fig. 1).

Molecular networking enabled meta-mass-shift chemical profiling¹⁷ (an analysis of chemical transformations on the basis of parent mass shifts between related spectra without the requirement of knowing the molecular structures) of the gastrointestinal tracts of germ-free and SPF mice. In colonized mice, there was a signature for water loss in the duodenum and jejunum and the loss of H₂, acetyl and methyl groups in later parts of the gastrointestinal tract (Fig. 1e). Of all the H₂ shifts, 23.1% were associated with bile acids, which indicates that colonization resulted in the oxidation of bile acids (a known microbial transformation)¹⁸. Deacetylations were also prevalent in colonized mice, although the metabolites on which this occurred remain unidentified. Germ-free mice had mass gains that corresponded to saccharides in all regions of the gastrointestinal tract (Fig. 1e); these gains were primarily associated with plant natural products, such as soyasaponins and flavonoids. The absence of these sugars in SPF mice implicates the microbiome in their metabolism (Extended Data Figs. 2, 3). A unique mass gain of C₄H₈ was detected in the jejunum and ileum of SPF mice (Fig. 1e) and 18.2% of spectra with this mass gain were derived from an unknown molecule related to the conjugated bile acid glycocholic acid (GCA) (Fig. 2a). Overall, both germ-free and SPF mice had frequent and diverse mass losses between related molecules, but in colonized mice there were fewer molecules that gained a molecular group (Fig. 1e). This indicates that the microbiome contributed more to the catabolic breakdown of molecules, and less to anabolism. However, we found the addition of C₄H₈ to GCA to be a particularly interesting anabolic reaction that was dependent on the gut microbiome, and we sought to investigate this further.

Discovery of new conjugated bile acids

Glycine- and taurine-conjugated bile acids were detected in both germ-free and SPF mice. The glycine and taurine amino acids were removed as they passed through the gastrointestinal tract in SPF mice only, which is a known microbial transformation¹⁹ (Fig. 1b, Extended Data Fig. 4). The molecular network of conjugated bile acids had several modified forms of these compounds that were present only in colonized mice, including the C₄H₈ addition that was related to the tandem mass spectra of GCA (Fig. 2a). Our analysis of the tandem mass spectra of three of these SPF-mouse nodes (*m/z* 556.363, 572.358 and 522.379) showed the maintenance of the core cholic acid, but with a fragmentation pattern that was characteristic of the presence of phenylalanine, tyrosine or leucine through an amide bond at the conjugation site in place of glycine or taurine (Extended Data Fig. 5, Supplementary Table 2). This represents a set of unique amino acid amide conjugations to cholic acid that are mediated by the microbiome, which create the newly identified bile acids phenylalanochoic acid (Phe-chol), tyrosocholic acid (Tyr-chol) and leucocholic acid (Leu-chol). These structures were validated with synthesized standards by retention time and MS/MS matching on several instrument platforms including targeted mass spectrometry (level one matches¹⁶) (Extended Data Figs. 5, 6, Supplementary Tables 2, 5). These molecules were detected in the duodenum, jejunum and ileum of SPF mice only, with tenfold-lower levels found in the caecum and colon after targeted mass spectrometry analysis using isotopically labelled internal standards (Supplementary Table 4). The liver-synthesized glycine and taurine conjugates were not only found in these same gut locations, but were also observed in the gall bladder and liver (Fig. 2b, Extended Data Fig. 6). Phe-chol was the most abundant microbial conjugate, on average, across the gastrointestinal tract; it was present at

147.0 nmol g⁻¹ tissue (s.d. ± 99.9) in the jejunum, 83.6 nmol g⁻¹ tissue (s.d. ± 81.3) in the ileum, 4.7 nmol g⁻¹ tissue (s.d. ± 3.4) in the caecum and 11.6 nmol g⁻¹ tissue (s.d. ± 12.2) in the colon. Phe-chol was present at its highest concentration at 447.2 nmol g⁻¹ tissue in a single sample from the jejunum (limit of detection (LOD) in Supplementary Tables 4, 6, 7).

The decreased abundance of these unique bile conjugates in the lower gastrointestinal tract prompted us to investigate whether there was reabsorption in the ileum or further metabolism by the microbiota. We collected portal and peripheral blood from an additional four SPF and six germ-free mice, and screened for the presence of conjugated bile acids. Both taurocholic acid and GCA were present in the portal and peripheral blood of colonized and sterile mice, but the newly identified amino acid amide conjugates were not detected (Extended Data Fig. 6). Furthermore, incubation of these molecules with an actively growing human faecal batch culture showed that the Tyr-, Phe- and Leu-conjugated bile acids were not deconjugated by the microbiota—even when deconjugation readily occurred on the host-synthesized GCA control, a well-known amidase hydrolase activity of bile acids that is mediated by the human microbiota²⁰ (Extended Data Fig. 6). However, oxidation of the cholate core occurred on all three of the newly identified conjugates, which indicates that they could be modified by microbial enzymes even when no concurrent oxidation of GCA was observed (Extended Data Fig. 6).

In the extensive literature relating to bile acids (comprising more than 42,000 publication records in PubMed^{21–27}), descriptions of unusual conjugations of bile acids are rare. Through 170 years of research into bile-acid chemistry, the accepted standard has been that mammalian bile acids are amide conjugated by a host liver enzyme (known as bile acid–CoA:amino acid *N*-acyltransferase (BAAT)) with either glycine or taurine. Here we report amide conjugations with phenylalanine, tyrosine and leucine associated with the microbiome in mice, and show that these compounds are common in humans.

Translation to humans

We performed a search using the Mass Spectrometry Search Tool (MASST) of 1,004 public datasets available in the Global Natural Products Social Molecular Networking (GNPS) database, which revealed spectral matches that correspond to Phe-chol, Tyr-chol and Leu-chol in 28 studies comprising samples from the gastrointestinal tract of both mice (3.2 to 59.4% of all samples) and humans¹¹ (1.6 to 25.3% of all samples) (Extended Data Fig. 7). In data from faecal samples collected for the American Gut Project²⁸, at least one of these unique bile acids was found in 1.6% of human faecal samples; Tyr-chol was the most prevalent (*n* = 490 samples) (Fig. 3a). These bile acids were found in higher frequency in samples from patients with inflammatory bowel disease or cystic fibrosis, or from infants, than in samples from the American gut project (Fig. 3a).

We reanalysed data deposited in the GNPS/MassIVE repository from a previously published study of the mouse microbiome and liver cancer, which enabled us to compare the abundance of the newly identified bile acid conjugates in mice fed a high-fat diet in comparison to their abundance when the mice were fed a normal chow with or without antibiotics²⁹ (Extended Data Fig. 7). The Phe, Tyr, and Leu amino acid conjugates were undetectable upon exposure to antibiotics, whereas GCA remained—supporting the role of the microbiome in the newly identified conjugation. In the same study²⁹, Phe-chol and Leu-chol were more abundant in mice fed a high-fat diet, with no change observed in the host-conjugated GCA (Extended Data Fig. 7). We further validated this association in data from a separate study in which atherosclerosis-prone mice fed a high-fat diet also had increased levels of the microbial conjugates, without a corresponding change in the host-produced taurocholic acid (Extended Data Fig. 7). Cystic fibrosis is known to result in insufficient production of pancreatic lipase, microbial dysbiosis and the build-up of fat in the gut³⁰. Reanalysis of the public data

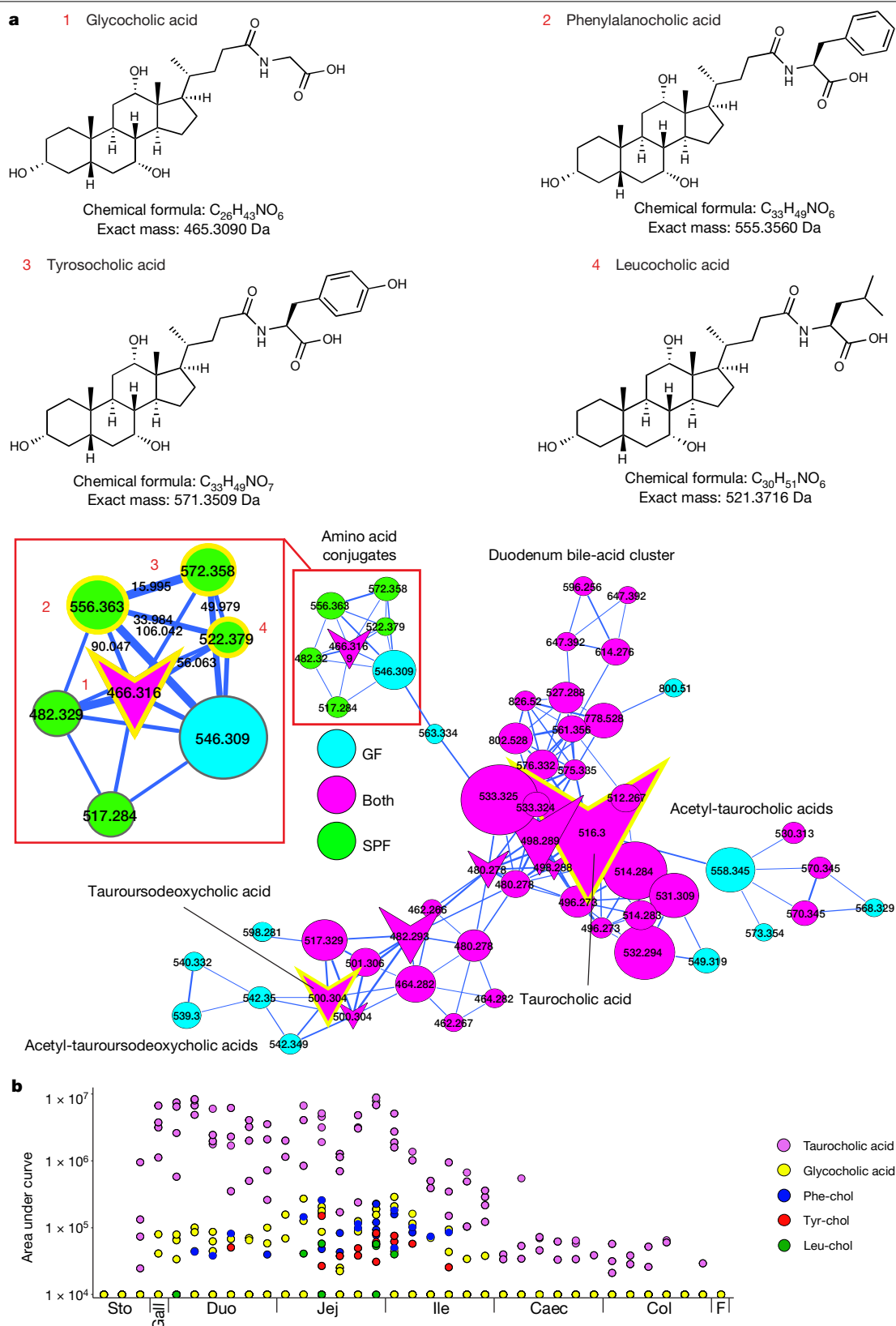


Fig. 2 | Newly identified microbial bile-acid conjugates. **a**, Structures and molecular networks of newly identified microbiome-conjugated bile acids, with host-conjugated GCA shown for comparison. The molecular network is coloured by mapping to germ-free or SPF mice (according to the colour legend). Inset highlights the parent masses and mass differences between the newly discovered molecules and GCA. Each node represents a clustered tandem mass spectrum; connections between the nodes indicate relationships

through the cosine score with their width scaled by the cosine size (cut-off minimum of 0.7). Circular nodes are unknown molecules, and arrowheads are spectra with matches in the GNPS libraries. **b**, Dot plot of the area-under-the-curve abundance of the newly identified and host-synthesized bile-acid conjugates in each SPF mouse ($n = 4$), through the mouse gastrointestinal tract and its subsections.

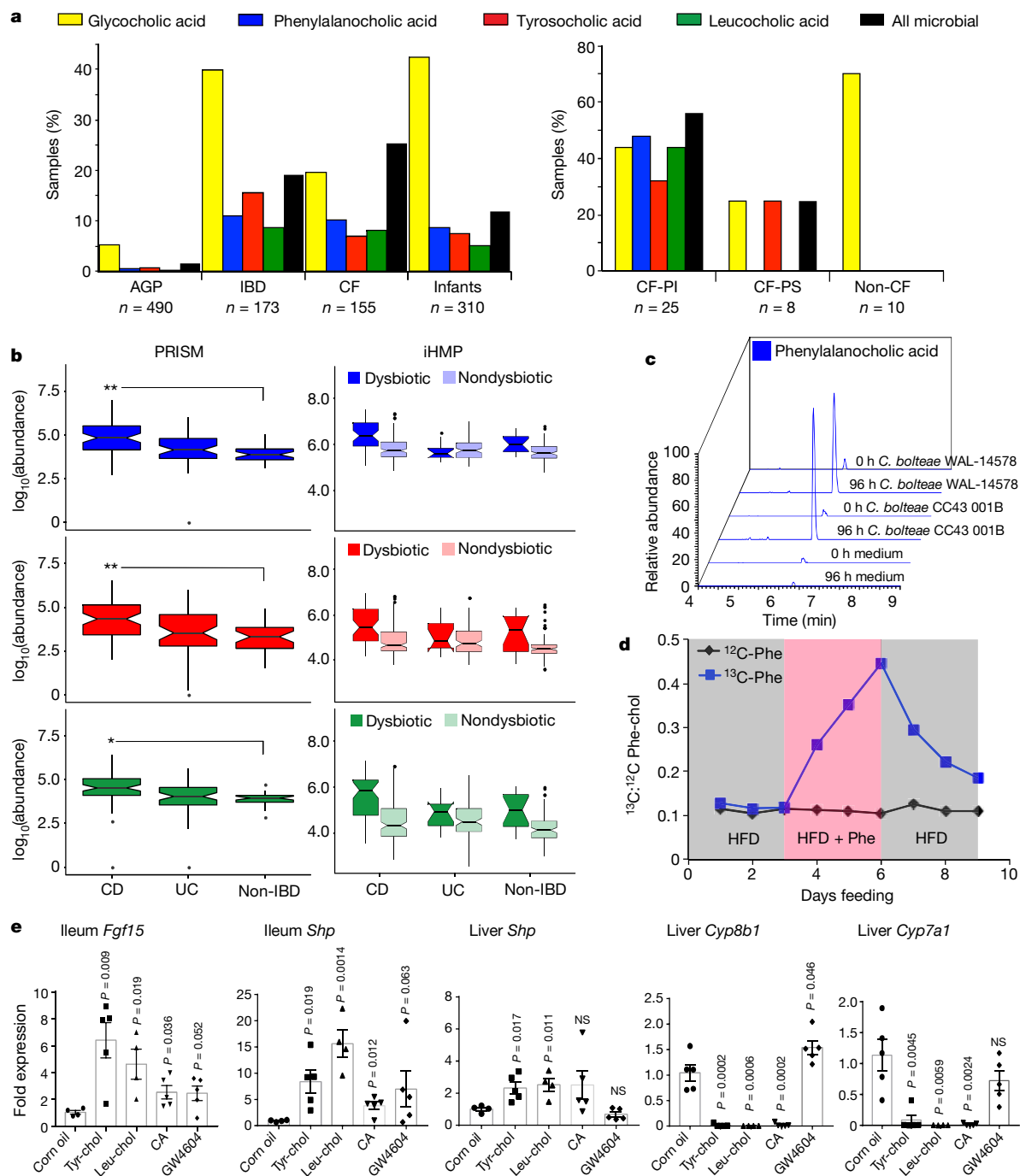


Fig. 3 | Presence, synthesis and function of microbial bile-acid conjugates.

a, Percentage of samples that were positive for the newly identified bile acids from GNPS public datasets and from paediatric patients with cystic fibrosis (compared to controls without cystic fibrosis). AGP, American gut project²⁸; CF, cystic fibrosis; IBD, inflammatory bowel disease; PS, pancreatic-sufficient; PI, pancreatic-insufficient. The colour coding of the bile acids applies to **a–c**. **b**, Abundance of the newly identified conjugates in the PRISM and iHMP (NIH Integrative Human Microbiome Project) datasets³¹. The statistical significance for the PRISM data was tested using the Wald's test (Crohn's disease (CD), $n = 68$ individuals; ulcerative colitis (UC), $n = 53$ individuals; noninflammatory bowel disease, $n = 34$ individuals) and for the iHMP dataset with a linear two-sided mixed-effects model. The iHMP comparisons are separated by type of inflammatory bowel disease, and by dysbiotic or nondysbiotic state (for ulcerative colitis, $n = 12$ dysbiotic and 110 nondysbiotic metabolomes; for Crohn's disease, $n = 48$ dysbiotic, and 169 nondysbiotic metabolomes; for noninflammatory bowel disease, $n = 15$ dysbiotic and 107 nondysbiotic metabolomes). Significance is shown using Benjamini–Hochberg-corrected

P values. Leu-chol, $q = 0.031$; Tyr-chol, $q = 0.0074$; Phe-chol, $q = 0.0043$.

* $q < 0.05$, ** $q < 0.01$. Boxes represent the interquartile range, notch is the 95% confidence interval of the mean, centre is the median and whiskers are 1.5× the interquartile range. **c**, Extracted ion chromatograms of Phe-chol from cultured isolates of *C. bolteae* compared to medium control at 0 h and 96 h (top). Experiment was performed twice. **d**, The ratio of ^{13}C -Phe-chol: ^{12}C -Phe-chol in faecal samples of a mouse fed a high-fat diet with ^{13}C -labelled phenylalanine (blue line) or unlabelled phenylalanine (black line) over time. Grey area indicates a three-day period during which a high-fat diet was fed; red indicates when the high-fat diet was supplemented with Phe. **e**, Quantitative PCR with reverse transcription data showing the mean and s.e.m. of the gene-expression ratio ($\Delta\Delta C_t$) of *Fgf15*, *Shp*, *Cyp7B1* and *Cyp7A1* to the *36B4* (also known as *Rplp0*) reference control in the ileum and/or liver of mice gavaged with different bile acids, compared to a mock control (corn oil) after 72 h. Statistical significance was tested against the mock control with a two-tailed t -test ($n = 4$ or 5 mice per group). CA, cholic acid.

from a cohort of paediatric patients, we found that these compounds were more prevalent in patients with cystic fibrosis (particularly in those with pancreatic insufficiency) than in healthy controls (Fig. 3a). Finally, detection of the newly identified conjugates in patients with inflammatory bowel disease led us to mine metabolome data from the second stage of the human microbiome project (HMP2)³¹, which focused on differences between controls and patients with inflammatory bowel disease, including patients with Crohn's disease or ulcerative colitis—subtypes of inflammatory bowel disease³¹ (Fig. 3b, Supplementary Table 8). All three metabolites were significantly higher in the dysbiotic state associated with patients with Crohn's disease, but not in patients with ulcerative colitis (Fig. 3b, Supplementary Data). Our MASST-based mining of public data from the GNPS database showed that these compounds are not only found in healthy humans but are also enriched in individuals with fatty guts and inflammatory bowel disease, which suggests that these compounds may have a potential role in (or be symptoms of) gut dysbiosis and human disease.

Microorganisms make the new bile acids

There was a strong positive correlation between the presence of a species of *Clostridium* and all three bile acids when mice were fed a high-fat diet (Pearson's r for Phe-chol, $r = 0.73$; for Tyr-chol, $r = 0.50$; and for Leu-chol, $r = 0.74$) (Extended Data Fig. 7, Supplementary Table 3). The clostridia are known to oxidize, epimerize and deconjugate bile acids^{32,33}. We therefore cultured 20 human gut microorganisms (with an emphasis on *Clostridium* species) in faecal culture medium³⁴ that contained amino acids and cholic acid precursors to screen for production of the newly identified conjugates. The *Clostridium bolteae* strains WAL-14578 and CC43001B both synthesized both Phe-chol and Tyr-chol (Extended Data Fig. 8). The addition of labelled ¹³C-phenylalanine to the medium verified that WAL-1457 could synthesize Phe-chol from the amino acid and cholate precursors (Extended Data Fig. 8). Similarly, we fed mice a high-fat diet with ¹³C-phenylalanine and were able to detect labelled Phe-chol in their faeces, which demonstrates microbial synthesis in vivo and shows that the amino acid precursors could come from the diet (Fig. 3d). *C. bolteae* is a bile-resistant gut bacterium that is more common in children with autism spectrum disorder³⁵, is associated with abdominal infections³⁶ and—together with *Blautia producta*—prevented colonization by vancomycin-resistant *Enterococcus* species in mice³⁷. The production of these bile acids by *C. bolteae* further verifies their association with the microbiota of the mouse gut, and implicates them as potentially important for intermicrobial interactions in the gut microbiome. However, addition of the newly identified conjugates to batch cultures of human faecal samples did not affect community structure (Extended Data Fig. 8), which led us to investigate how these compounds may affect gut physiology through host receptor signalling.

New bile acids and the farnesoid X receptor

The farnesoid X receptor (FXR) is a key receptor for bile acids that is expressed in the intestine, liver and other tissues. The most-potent naturally occurring agonistic ligand of FXR is chenodeoxycholic acid, whereas tauro- β -muricholic acid is an FXR antagonist³⁸. To assess the ability of the newly identified bile acids to affect human FXR signalling, we established a luciferase reporter assay in human embryonic kidney (HEK)293 cells³⁹. Phe-chol and Tyr-chol were strong human-FXR agonists (Extended Data Fig. 9, Supplementary Table 9). The phenylalanine conjugate ($R^2 = 0.92$, half maximal effective concentration (EC_{50}) = 4.5 μ M) was twice as strong of an agonist as chenodeoxycholic acid ($R^2 = 0.88$, $EC_{50} = 9.7 \mu$ M), and the tyrosine conjugate was the most potent of them all ($R^2 = 0.93$, $EC_{50} = 0.14 \mu$ M). Furthermore, gavage of mice with these compounds increased expression of the FXR effector genes *Fgf15* and *Shp* (also known as *NrOb2*) in the intestine (12.2- and

13.3-fold with Tyr-chol at 24 h, $P = 0.029$ and 0.009 ; 6.2 and 9.3-fold at 72 h, $P = 0.009$ and 0.019) (Fig. 3e, Extended Data Fig. 9). Although *Shp* expression did not change detectably in the liver at 24 h after gavage, levels were increased 2.3-fold after 72 h ($P = 0.017$) (Fig. 3e, Extended Data Fig. 9). Changes in expression of the bile-acid synthesis genes *Cyp7a1* and *Cyp8b1* also showed a time-dependent effect. *Cyp7a1* was at 9% of control levels at 24 h ($P = 0.001$) and *Cyp8b1* was at 69% ($P = 0.004$) (Extended Data Fig. 9). At 72 h (after 4 gavages), *Cyp7a1* expression was at 8% of control levels ($P = 0.004$), and for *Cyp8b1* the transcript was further reduced to 2% ($P = 0.0002$) (Fig. 3e). The strong time-dependent reduction in liver *Cyp7a1* and *Cyp8b1* transcripts indicates that—similar to the primary bile acid cholic acid—gavage of mice with the newly identified compounds reduced the expression of downstream FXR-target genes that are responsible for bile-acid synthesis in the liver. However, the possibility that this effect was due to FXR agonism through release of cholate from amide conjugate hydrolysis cannot be excluded.

Bile-acid metabolism by the microbiome was first described in the 1960s⁴⁰. The four known mechanisms of microbial metabolism are dehydroxylation, dehydration and epimerization of the cholesterol backbone, and deconjugation of the amino acids glycine or taurine^{1,41,42}. Here, we identify bile-acid transformation by the microbiome mediated by a fifth and completely different mechanism: amide conjugation of the cholate backbone with the amino acids phenylalanine, tyrosine and leucine. Although there are homologues of the human bile-acid-conjugation gene *BAAT* in clostridial genomes, the microbial enzyme in question remains unknown. Regardless of the mechanism of their synthesis, the newly identified conjugates stimulate the human FXR receptor in a cell-based system and the expression of FXR-target genes that are responsible for bile-acid production in the liver were reduced when administered to mice. Additional studies are needed to understand the health implications of bile-acid reconjugation by the human microbiome and its potential effects on FXR-related diseases.

Conclusion

This study shows that the chemistry of all organ systems is affected by the presence of the microbiome. The strongest signatures come from the gut, particularly via the breakdown of plant natural products from food and the manipulation of bile acids. The microbiome is primarily a catabolic entity, breaking down compounds through the enzymatic removal of chemical groups. However, we found an anabolic reaction that represents a fifth mechanism of bile-acid metabolism by the microbiome, which operates through unique amino acid conjugations of cholic acid. As the connections between humans and our microbial symbionts become increasingly appreciated, a combination of globally untargeted approaches and the development of tools that interlink these datasets (such as the GNPS and MASST analysis infrastructure) will enable the more-efficient characterization of microbial molecules and efficient translation between model animals and human studies, leading to a better understanding of the deep connection between our microbiota, our metabolites and our health.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2047-9>.

- Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. Bile acids and the gut microbiome. *Curr. Opin. Gastroenterol.* **30**, 332–338 (2014).
- Gilbert, J. A. et al. Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94–103 (2016).
- Wikoff, W. R. et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl Acad. Sci. USA* **106**, 3698–3703 (2009).

4. Marcobal, A. et al. Metabolome progression during early gut microbial colonization of gnotobiotic mice. *Sci. Rep.* **5**, 11589 (2015).
5. Miller, T. L. & Wolin, M. J. Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Appl. Environ. Microbiol.* **62**, 1589–1592 (1996).
6. Gillner, M., Bergman, J., Cambillau, C., Fernström, B. & Gustafsson, J. A. Interactions of indoles with specific binding sites for 2,3,7,8-tetrachlorodibenzo-p-dioxin in rat liver. *Mol. Pharmacol.* **28**, 357–363 (1985).
7. Martin, F.-P. J. et al. A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Mol. Syst. Biol.* **3**, 112 (2007).
8. Moriya, T., Satomi, Y., Murata, S., Sawada, H. & Kobayashi, H. Effect of gut microbiota on host whole metabolome. *Metabolomics* **13**, 101 (2017).
9. Swann, J. R. et al. Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc. Natl Acad. Sci. USA* **108** (Suppl 1), 4523–4530 (2011).
10. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
11. Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
12. Watrous, J. et al. Mass spectral molecular networking of living microbial colonies. *Proc. Natl Acad. Sci. USA* **109**, E1743–E1752 (2012).
13. Protsyuk, I. et al. 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. *Nat. Protocols* **13**, 134–154 (2018).
14. Hofmann, A. F. & Hagey, L. R. Key discoveries in bile acid chemistry and biology and their clinical applications: history of the last eight decades. *J. Lipid Res.* **55**, 1553–1595 (2014).
15. Yang, J. Y. et al. Molecular networking as a dereplication strategy. *J. Nat. Prod.* **76**, 1686–1699 (2013).
16. Sumner, L. W. et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
17. Hartmann, A. C. et al. Meta-mass shift chemical profiling of metabolomes from coral reefs. *Proc. Natl Acad. Sci. USA* **114**, 11685–11690 (2017).
18. Hirano, S. & Masuda, N. Characterization of NADP-dependent 7 β -hydroxysteroid dehydrogenases from *Peptostreptococcus productus* and *Eubacterium aerofaciens*. *Appl. Environ. Microbiol.* **43**, 1057–1063 (1982).
19. Wahlström, A., Sayin, S. I., Marschall, H.-U. & Bäckhed, F. Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metab.* **24**, 41–50 (2016).
20. Huijghebaert, S. M. & Hofmann, A. F. Influence of the amino acid moiety on deconjugation of bile acid amides by cholyglycine hydrolase or human fecal cultures. *J. Lipid Res.* **27**, 742–752 (1986).
21. Myher, J. J., Marai, L., Kuksis, A., Yousef, I. M. & Fisher, M. M. Identification of ornithine and arginine conjugates of cholic acid by mass spectrometry. *Can. J. Biochem.* **53**, 583–590 (1975).
22. Peric-Golia, L. & Jones, R. S. Ornithocholanic acids and cholelithiasis in man. *Science* **142**, 245–246 (1963).
23. Gordon, B. A., Kuksis, A. & Beveridge, J. M. R. Separation of bile acid conjugates by ion exchange chromatography. *Can. J. Biochem. Physiol.* **41**, 77–89 (1963).
24. Yousef, I. M. & Fisher, M. M. Bile acid metabolism in mammals. VIII. Biliary secretion of cholyarginine by the isolated perfused rat liver. *Can. J. Physiol. Pharmacol.* **53**, 880–887 (1975).
25. Tamari, M., Ogawa, M. & Kametaka, M. A new bile acid conjugate, ciliatocholic acid, from bovine gall bladder bile. *J. Biochem.* **80**, 371–377 (1976).
26. Hagey, L. R., Schteingart, C. D., Rossi, S. S., Ton-Nu, H. T. & Hofmann, A. F. An N-acyl glycytaurine conjugate of deoxycholic acid in the biliary bile acids of the rabbit. *J. Lipid Res.* **39**, 2119–2124 (1998).
27. Nair, P. P., Solomon, R., Bankoski, J. & Plapinger, R. Bile acids in tissues: binding of lithocholic acid to protein. *Lipids* **13**, 966–970 (1978).
28. McDonald, D. et al. American gut: an open platform for citizen science microbiome research. *mSystems* **3**, e00031-18 (2018).
29. Shalapour, S. et al. Inflammation-induced IgA⁺ cells dismantle anti-liver cancer immunity. *Nature* **551**, 340–345 (2017).
30. Manor, O. et al. Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis. *Sci. Rep.* **6**, 22493 (2016).
31. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).
32. Hirano, S., Masuda, N., Oda, H. & Mukai, H. Transformation of bile acids by *Clostridium perfringens*. *Appl. Environ. Microbiol.* **42**, 394–399 (1981).
33. Winston, J. A. & Theriot, C. M. Impact of microbial derived secondary bile acids on colonization resistance against *Clostridium difficile* in the gastrointestinal tract. *Anaerobe* **41**, 44–50 (2016).
34. McDonald, J. A. K. et al. Evaluation of microbial community reproducibility, stability and composition in a human distal gut chemostat model. *J. Microbiol. Methods* **95**, 167–174 (2013).
35. Finegold, S. M. et al. Pyrosequencing study of fecal microflora of autistic and control children. *Anaerobe* **16**, 444–453 (2010).
36. Dehoux, P. et al. Comparative genomics of *Clostridium bolteae* and *Clostridium clostridioforme* reveals species-specific genomic properties and numerous putative antibiotic resistance determinants. *BMC Genomics* **17**, 819 (2016).
37. Caballero, S. et al. Cooperating commensals restore colonization resistance to vancomycin-resistant *Enterococcus faecium*. *Cell Host Microbe* **21**, 592–602.e4 (2017).
38. Sayin, S. I. et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab.* **17**, 225–235 (2013).
39. Downes, M. et al. A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. *Mol. Cell* **11**, 1079–1092 (2003).
40. Gustafsson, B. E., Gustafsson, J. A. & Sjövall, J. Intestinal and fecal sterols in germfree and conventional rats. Bile acids and sterols 172. *Acta Chem. Scand.* **20**, 1827–1835 (1966).
41. Midtvedt, T. Microbial bile acid transformation. *Am. J. Clin. Nutr.* **27**, 1341–1347 (1974).
42. Gérard, P. Metabolism of cholesterol and bile acids by the gut microbiota. *Pathogens* **3**, 14–24 (2013).
43. Wang, M. et al. Mass spectrometry searches using MASST. *Nat. Biotechnol.* **38**, 23–26 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Article

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All metabolomics data that support the findings of this study are available at GNPS (<https://gnps.ucsd.edu/>) under MassIVE ID numbers: MSV000079949 (original germ-free and SPF mouse data), MSV000082480, MSV000082467, MSV000079134, MSV000082406, MSV000083032, MSV000083004 and MSV000083446. The sequencing data for the germ-free and SPF mouse study are available on the Qiita microbiome data analysis platform at <https://qiita.ucsd.edu/> under study ID 10801 and through the European Bioinformatics Institute accession number ERP109688. Source Data for Figs. 1–3, Extended Data Fig. 7 are provided with the paper.

Code availability

MASST can be accessed at <https://masst.ucsd.edu/>; the development of MASST is described in ref.⁴³. The code for MS/MS-based MASST searching is available at https://github.com/CCMS-UCSD/GNPS_Workflows/tree/master/search_single_spectrum.

Acknowledgements The authors acknowledge funding from the National Institutes of Health (NIH), grants 5U01AI124316-03, 1R03CA211211-01, 1R01HL116235 U54DE023798, R24DK110499, GMS10RRO29121, 1 DP1 AT010885, P30 DK120515 and R01HD084163. Additionally, B.S.B. was supported by UCSD KL2 (1KL2TR001444), T.D.A. by the National Library of Medicine Training Grant NIH grant T15LM011271. R.M.E. is an investigator of the Howard Hughes Medical Institute and March of Dimes Chair in Molecular and Developmental Biology at the Salk Institute. R.M.E. was funded by grants from the NIH (DK057978, HL105278, HL088093 and ES010337), and Samuel Waxman Cancer Research Foundation. We acknowledge G. Ackermann for her contributions. This work was also supported in part by Seed Grants from the UC San Diego Center for Microbiome Innovation. This work was funded by grants from the NIH (DK057978, HL105278 and HL088093), National Cancer Institute (CA014195), the Leona M. and Harry B. Helmsley Charitable Trust (2017PG-MED001), SWCRF Investigator Award and Ipsen/Biomeasure. J.L. is supported by grant EIA14660045, an American Heart Association

Established Investigator Award. T.F. is supported by a Hewitt Medical Foundation Fellowship, a Salk Alumni Fellowship. T.F., R.K. and P.C.D. acknowledge support from the Crohn's & Colitis Foundation (CCFA). R.M.E. and M.D. are supported in part by a Stand Up to Cancer (SU2C) - Cancer Research UK-Lustgarten Foundation Pancreatic Cancer Dream Team Research Grant (SU2C-AACR-DT-20-16). SU2C is a programme of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C. Research reported in this publication was also supported by the National Institute of Environmental Health Sciences of the NIH under Award Number P42ES010337. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions P.C.D., R.K. and R.A.Q. designed the project. P.C.D. and R.A.Q. discovered the bile acids. R.A.Q., A.A., A.V.M., F.V., J.M.G., N.G., A.T., M.P.C., L.K.C., A.T.N., M.J.M., G.H., M.P., C.C., S.B., E.G., K.C.W., P.B.-F., H.M., E.A.F., H.V., J.L.-P., T.D.A., A.C.K., J.A.-P., Z.B., E.L., M.Q., T.F., R.D.W., D.F., G.G.H. and R.B. generated data. R.A.Q., A.V., A.T., Z.B., A.V.M., R.d.S., R.J.X., T.F., R.D.W., M.D., R.M.E. and M.P.C. analysed data. R.A.Q., B.S.B., M.L., O.P., J.T.C., M.L., J.C.L., K.A.P., B.I.K., R.J., M.E., K.E.R., G.H., C.C., W.J.S. and R.B. collected samples. D.S., E.G., M.P.C. and A.T.N. were responsible for chemical synthesis. T.F., R.D.W., M.D. and R.M.E. were responsible for the FXR in vitro and in vivo functional studies and analysis, and edited the paper. P.C.D., R.K., S.K.M., V.N., C.H., R.J.X., A.C.K. and D.S. guided experimental design and analysis. M.W. converted the data in GNPS and developed MASST spectral search. T.T., V.N., M.R. and S.K.M. raised mice and guided experimental design. R.A.Q. and P.C.D. wrote the manuscript.

Competing interests C.H. is on the scientific advisory board of Seres Therapeutics. M.W. is founder of, and A.A. is a consultant for, Omata Laboratories LLC. P.C.D. and M.W. are consultants for Sirenas Therapeutics. W.J.S. consults for Abbvie, Allergan, Amgen, Arena Pharmaceuticals, Avexgen Therapeutics, BeiGene, Boehringer Ingelheim, Celgene, Celltrion, Conatus, Cosmo, Escalier Biosciences, Ferring, Forbion, Genentech, Gilead Sciences, Gossamer Bio, Incyte, Janssen, Kyowa Kirin Pharmaceutical Research, Landos Biopharma, Lilly, Oppilan Pharma, Otsuka, Pfizer, Progenity, Prometheus Biosciences (merger of Precision IBD and Prometheus Laboratories), Reistone, Ritter Pharmaceuticals, Robarts Clinical Trials (owned by Health Academic Research Trust, HART), Series Therapeutics, Shire, Sienna Biopharmaceuticals, Sigmoid Biotechnologies, Sterna Biologicals, Sublimity Therapeutics, Takeda, Theravance Biopharma, Tigenix, Tillotts Pharma, UCB Pharma, Ventyx Biosciences, Vimalan Biosciences and Vivelix Pharmaceuticals; and holds stock or stock options from BeiGene, Escalier Biosciences, Gossamer Bio, Oppilan Pharma, Prometheus Biosciences (merger of Precision IBD and Prometheus Laboratories), Progenity, Ritter Pharmaceuticals, Ventyx Biosciences and Vimalan Biosciences.

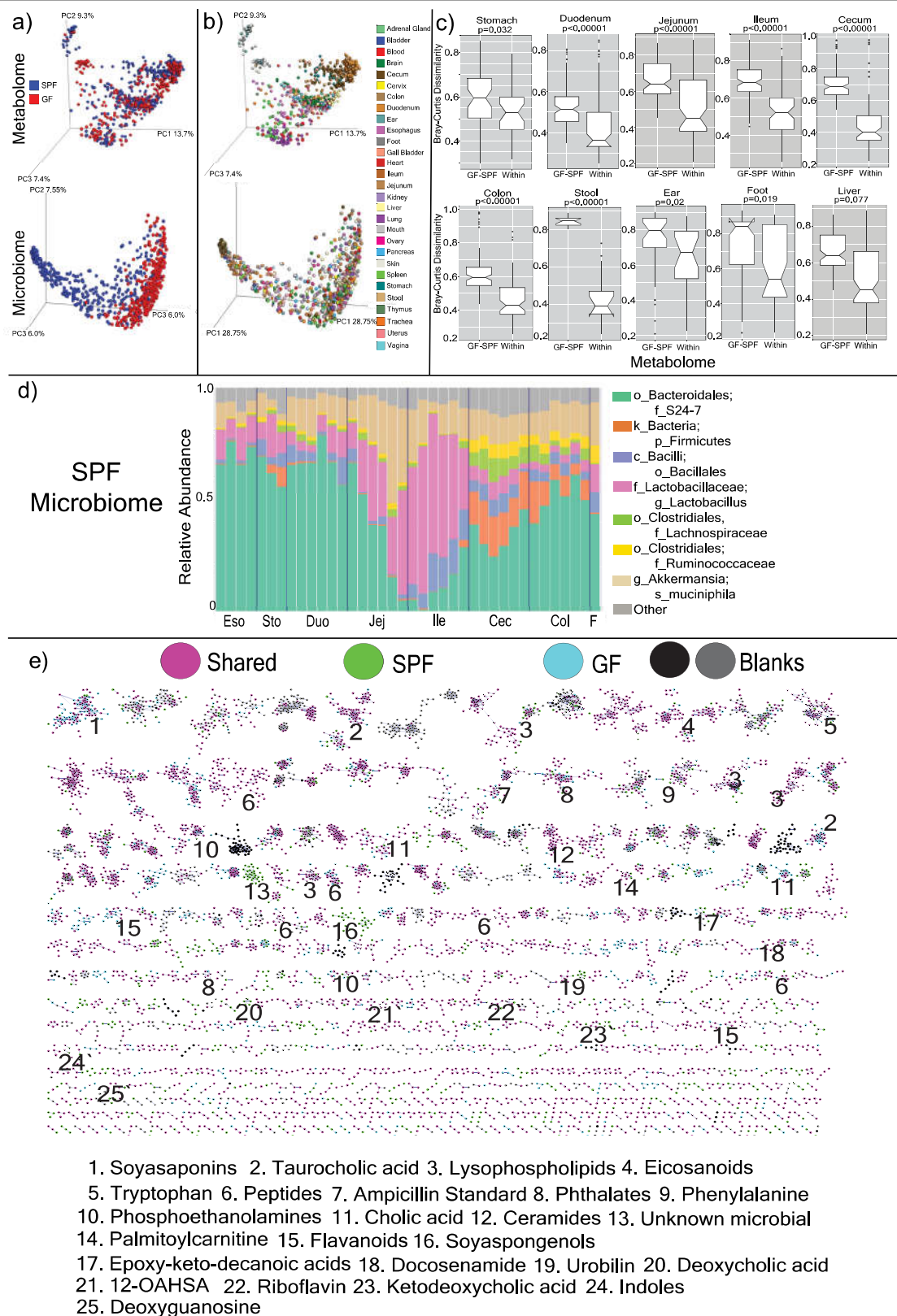
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2047-9>.

Correspondence and requests for materials should be addressed to P.C.D.

Peer review information *Nature* thanks Hanns-Ulrich Marschall, Trent Northen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

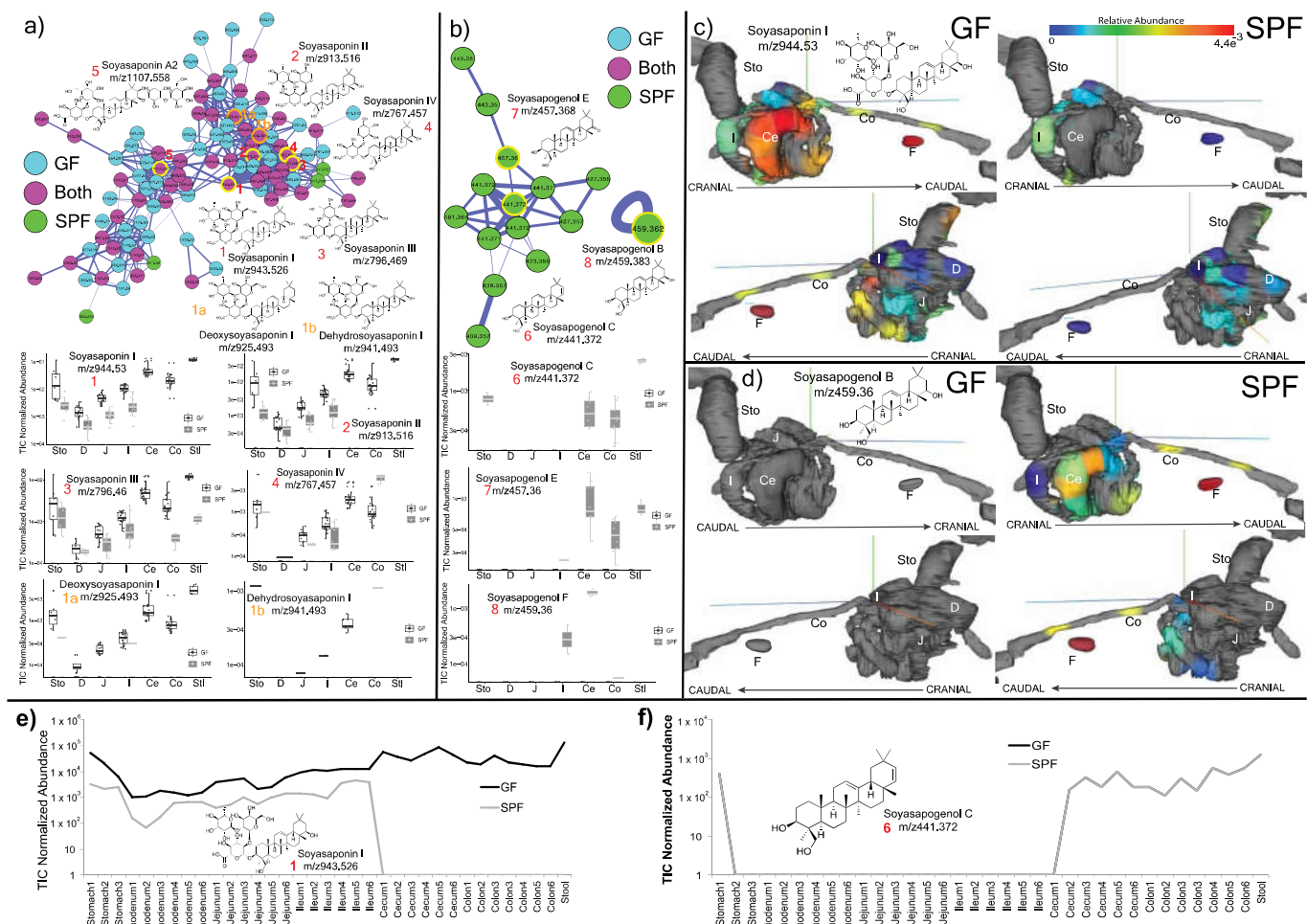
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Microbiome and metabolome diversity in germ-free and SPF mice. **a**, Principal coordinate (PC) analysis of microbiome and mass-spectrometry data highlighted by sample source as germ-free (GF) or SPF ($n = 4$ mice in each group). The microbial signatures from the germ-free mice are an important control, which represents background reads found in buffers, tips and tubes and other experimental materials. **b**, Data from **a** highlighted by organ source ($n = 4$ mice in each group). **c**, Bray–Curtis dissimilarities of the metabolome data collected from mouse organs. The dissimilarities are calculated within individual mice of the same group (germ-free or SPF, ‘within’) or across the germ-free and SPF groups (‘GF-SPF’) ($n = 4$ mice in each group). Only samples collected from exact same location (subsection) are compared. Significance was tested with a two-sided Mann–Whitney U -test. Boxes

represent the interquartile range (IQR), the notch is the 95% confidence interval of the mean, the centre is the median and whiskers are $1.5 \times$ the IQR. **d**, Microbiome profile of the gastrointestinal tracts of SPF mice. Data were generated by sequencing 16S rRNA gene amplicons from each organ and organ section, and analysed through the Qiita Deblur pipeline as described in the Supplementary Methods. Bacterial taxa of relevance are colour-coded according to the legend. **e**, Molecular network of LC–MS/MS data with nodes coloured by source as germ-free, SPF, shared or detected in blanks. Molecular families with metabolites annotated by spectral matching in GNPS are listed by a number that corresponds to the molecular family. These are level-2 or -3 annotations according to the metabolomics standards consortium¹⁶. 12-OAHSA, 12-(9Z-octadecenoyloxy)-octadecanoic acid.

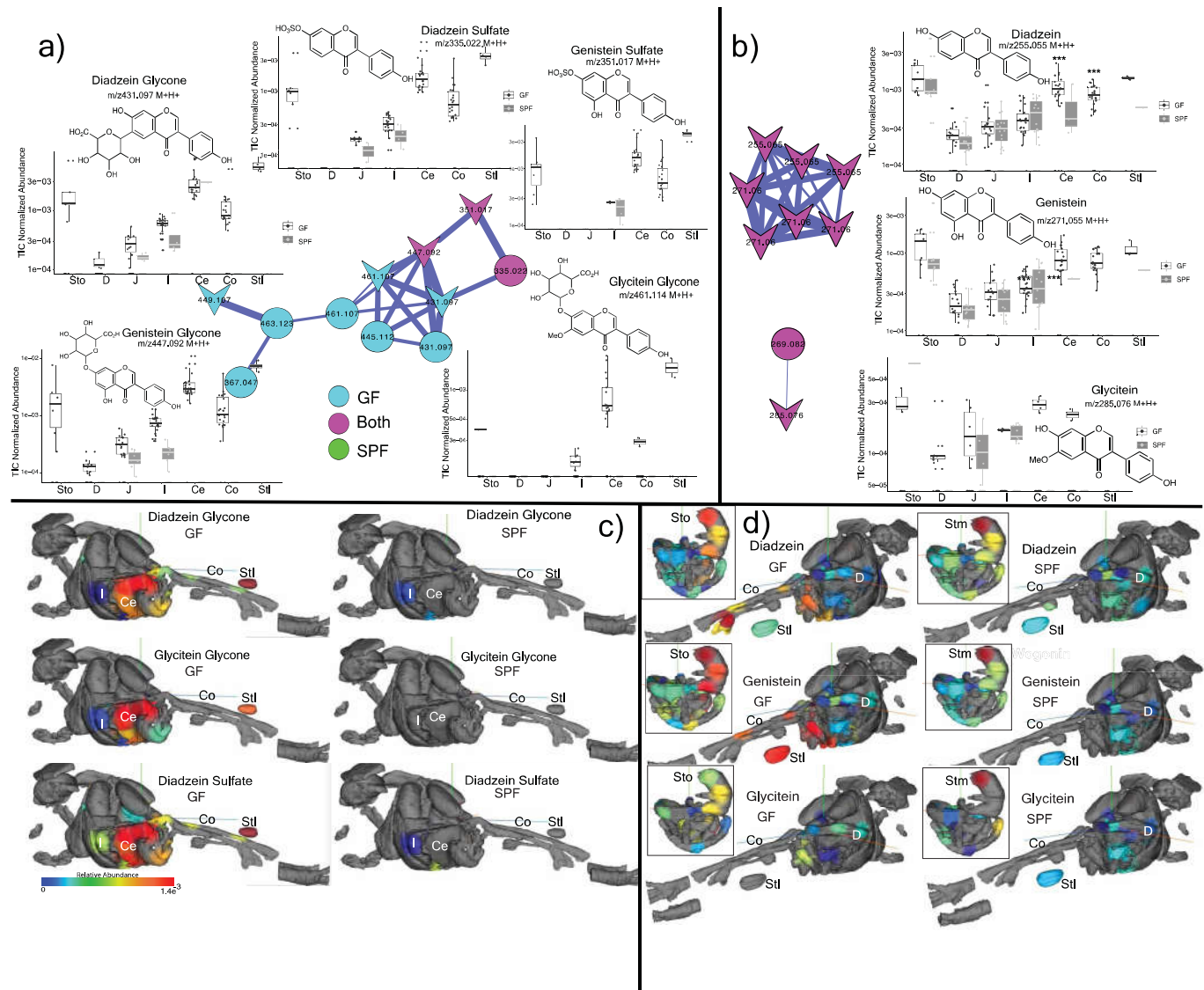


Extended Data Fig. 2 | Microbial metabolism of soyasaponins in metabolomics data from germ-free and SPF mice. $n = 4$ mice in each group.

a, Molecular network cluster of soyasaponins, coloured by source of each node as germ-free, SPF or shared. Structures of corresponding molecules are shown in nodes highlighted in yellow, according to the numbering scheme. Mean total-ion-current-normalized abundance of each soyasaponin metabolite from the gastrointestinal tracts of germ-free and SPF mice. Ce, caecum; co, colon; D, duodenum; I, ileum; J, jejunum; stl, stool; sto, stomach. Boxes represent the IQR, the centre is the median and whiskers are $1.5 \times$ the IQR. $n = 4$ mice in each group.

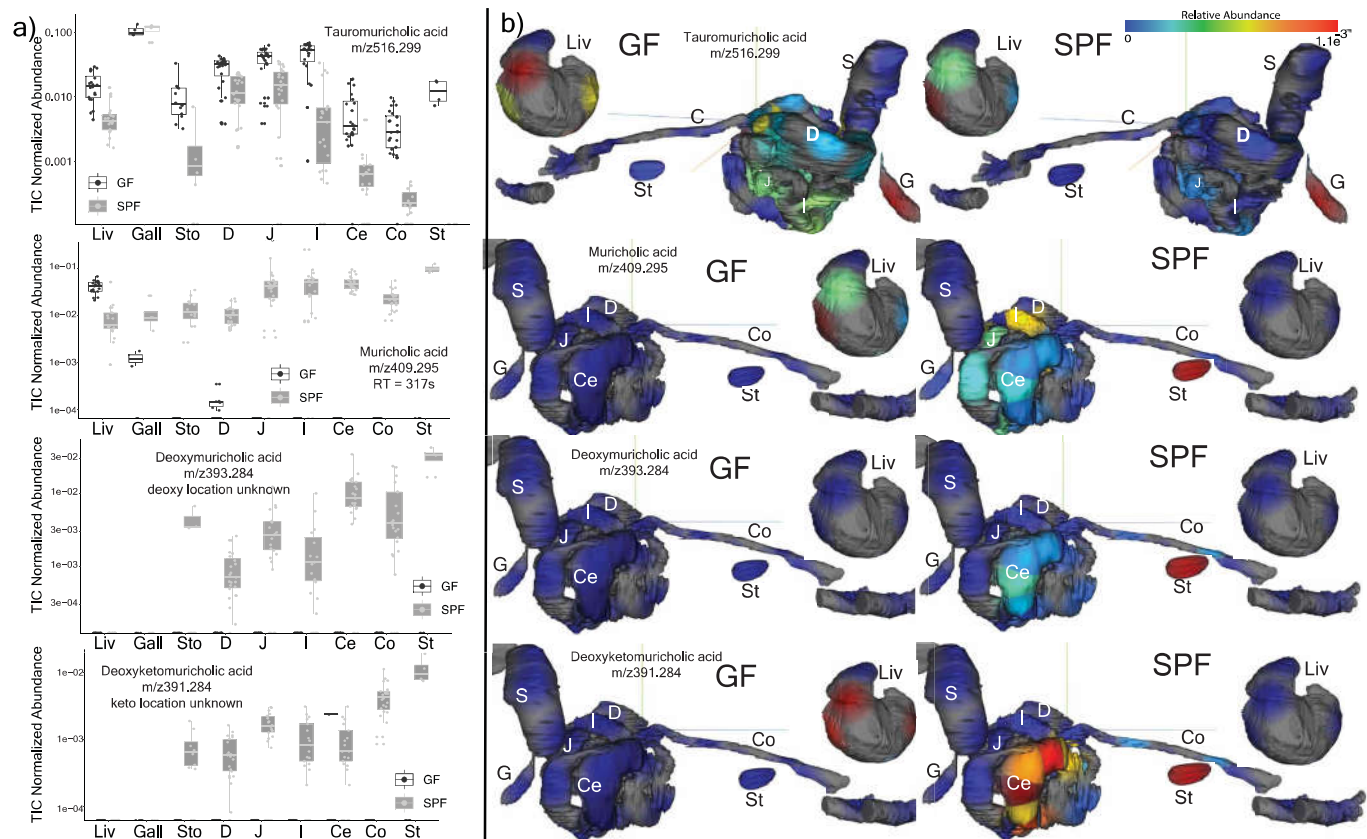
b, Molecular family of soyasapogenols, their structures and relative abundances in gut organs of germ-free and SPF mice (data are in the same

format as in a). **c,** Three-dimensional model visualization (generated using 'ili') of the normalized abundance of soyasaponin I in the mouse gastrointestinal tract. The abundance of the metabolite is indicated according to the rainbow spectrum (high, red; low, blue). $n = 4$ mice in each group. **d,** Three-dimensional cartography (generated using 'ili') of the normalized abundance of soyasapogenol B onto an magnetic resonance imaging organ model of the mice. **e,** Mean normalized abundance of soyasaponin I through all gastrointestinal sample locations in the germ-free and SPF mice. **f,** Mean normalized abundance of soyasapogenol through all gastrointestinal sample locations. The annotations are level two or three¹⁶.

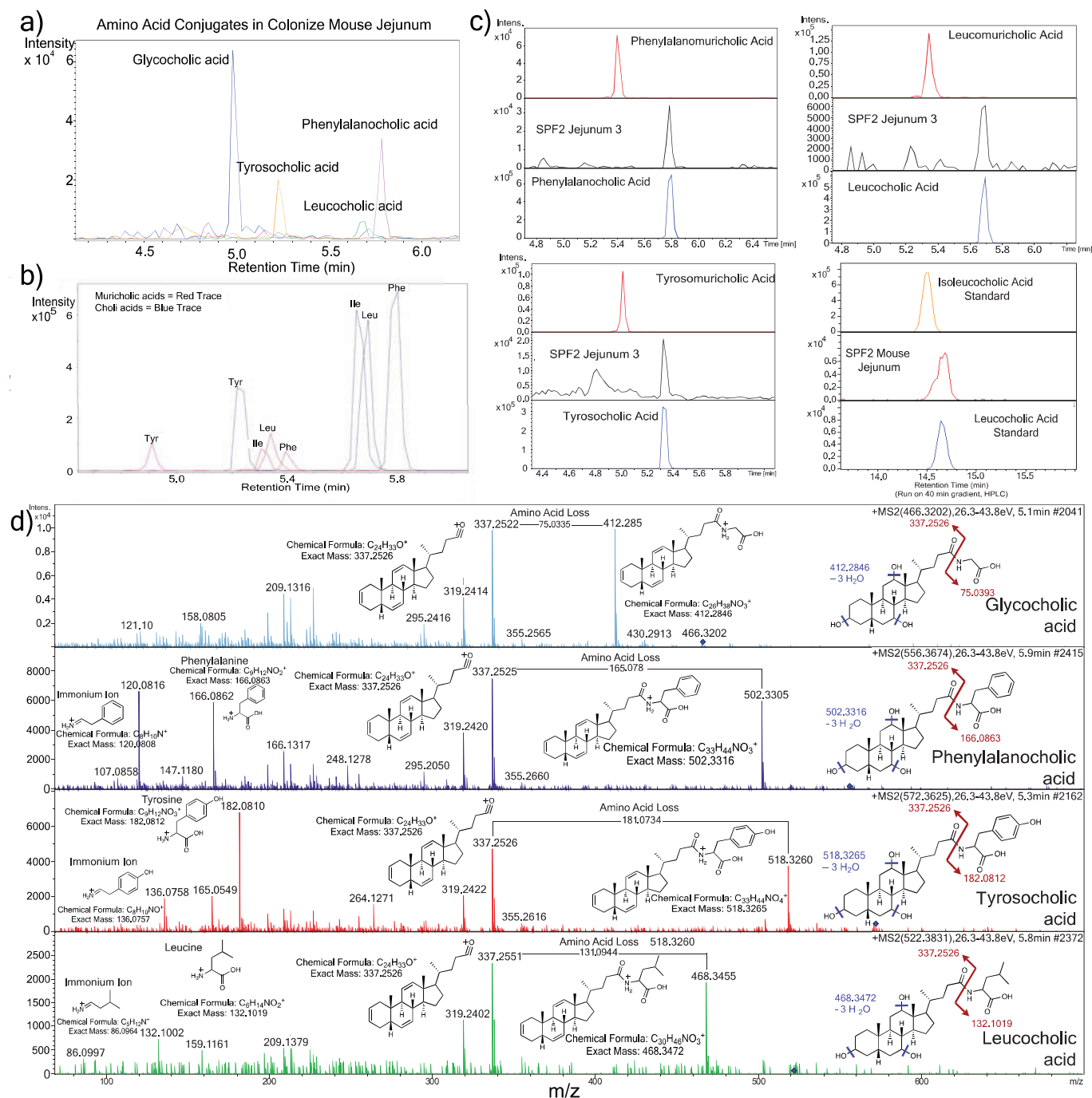


Extended Data Fig. 3 | Microbial metabolism of plant isoflavones in metabolomics data from germ-free and SPF mice. a, Structures, molecular network and total-ion-chromatogram-normalized abundance of glycone isoflavonoids in the mouse gastrointestinal tract. Nodes are coloured according to their source in germ-free or SPF mice ($n = 4$ mice each), and known library hits are shaped as arrowheads. Boxes represent the IQR, the centre is the median and whiskers are 1.5x the IQR. **b,** Same information as in **a**, for the

aglycones. **c,** Three-dimensional molecular cartography mapping the abundance of the daidzein and glycitein glycone and sulfate forms through entire 3D mouse model. The normalized abundance of a particular molecule is indicated as a heat map. Red, most abundant; blue, least abundant. **d,** Three-dimensional molecular cartography mapping the abundance of the daidzein and glycitein aglycone forms through entire 3D mouse model. The gastrointestinal tract model is inset for reference. The annotations are level two or three¹⁶.



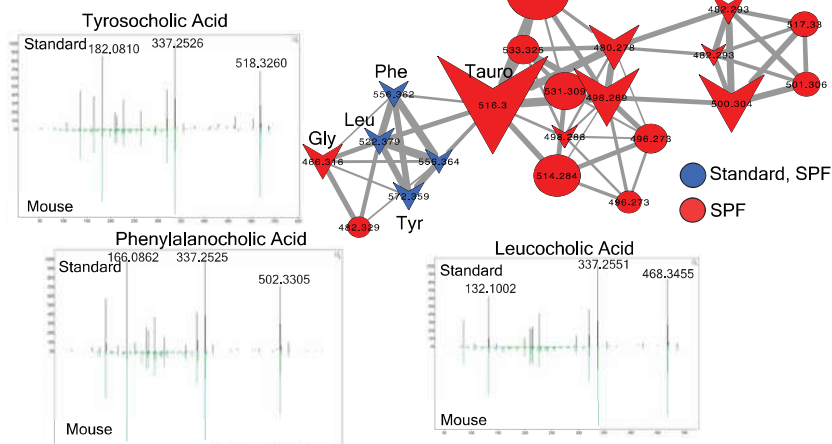
cartography mapping the abundance of the same bile acids as in **a** through the mouse gastrointestinal-tract model; liver is separated for better visualization. The normalized abundance of a particular molecule is indicated as a heat map. Red, most abundant; blue, least abundance. The annotations are level two or three¹⁶.



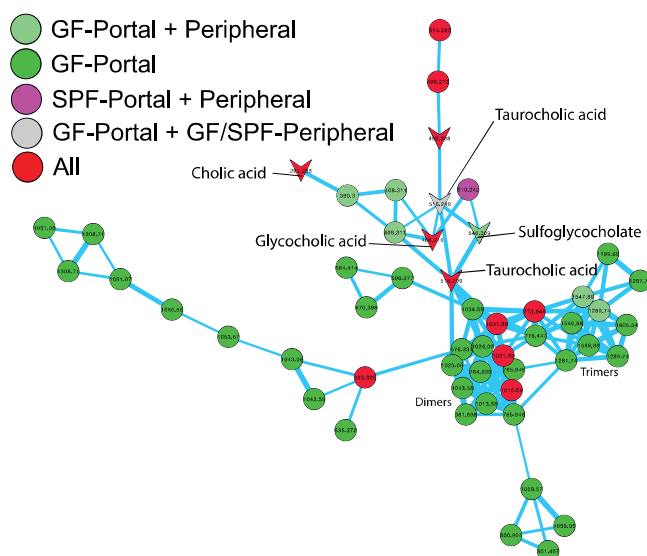
Extended Data Fig. 5 | Mass spectrometry analysis of newly identified conjugated bile acids. a, Extracted-ion-chromatogram MS¹ traces of Tyr-chol (m/z 572.37 \pm 0.05 Da), Phe-chol (m/z 556.37 \pm 0.05 Da) and Leu-chol (m/z 522.37 \pm 0.05 Da). Experiments were performed four times. **b,** Extracted ion chromatograms for the synthetic muricholic and cholic acid versions of the Phe (m/z 556.37 \pm 0.05), Tyr (572.37 \pm 0.05) and Leu (522.37 \pm 0.05) conjugates, showing the different retention times from the muricholic- and cholic-acid forms. **c,** Retention time alignments of synthetic muricholic- and cholic-acid conjugates with the newly identified conjugates found in a sample from the

jejunum of a colonized mouse. The isoleucocholic- and leucocholic-acid analysis was run on a long-gradient high-performance liquid-chromatography column to separate isomeric Ile and Leu conjugates, and to compare to those detected *in vivo*. **d,** Annotation of MS/MS fragmentation patterns for the three conjugated bile acids and GCA. Structures of the immonium ions from amino acid fragmentation, whole amino acid fragments and the major sterol fragment are shown. Loss of the amino acid mass on the bile-acid steroid backbone is also highlighted.

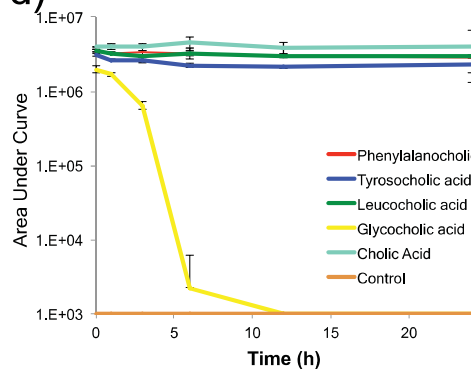
a)



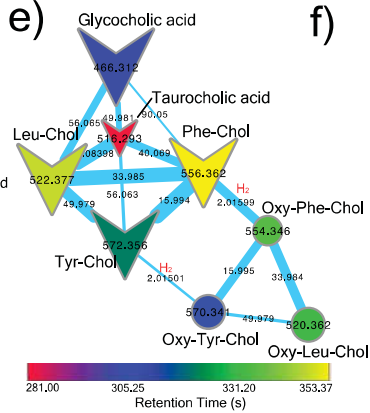
c)



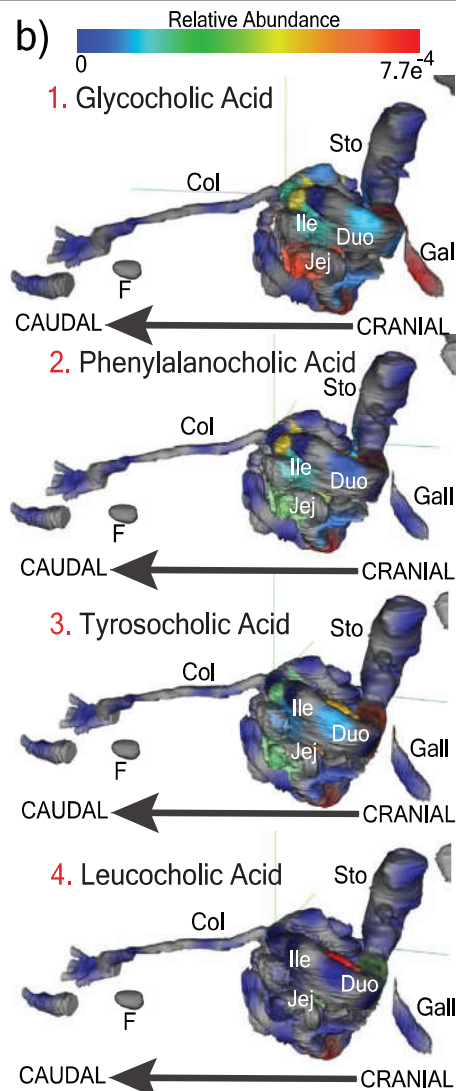
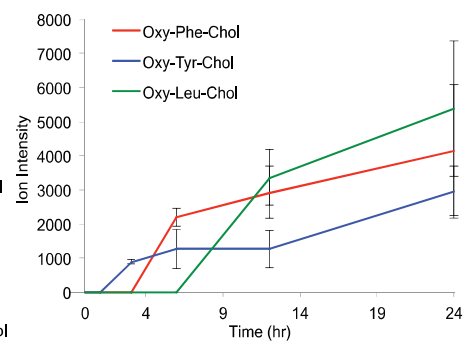
d)



e)



f)

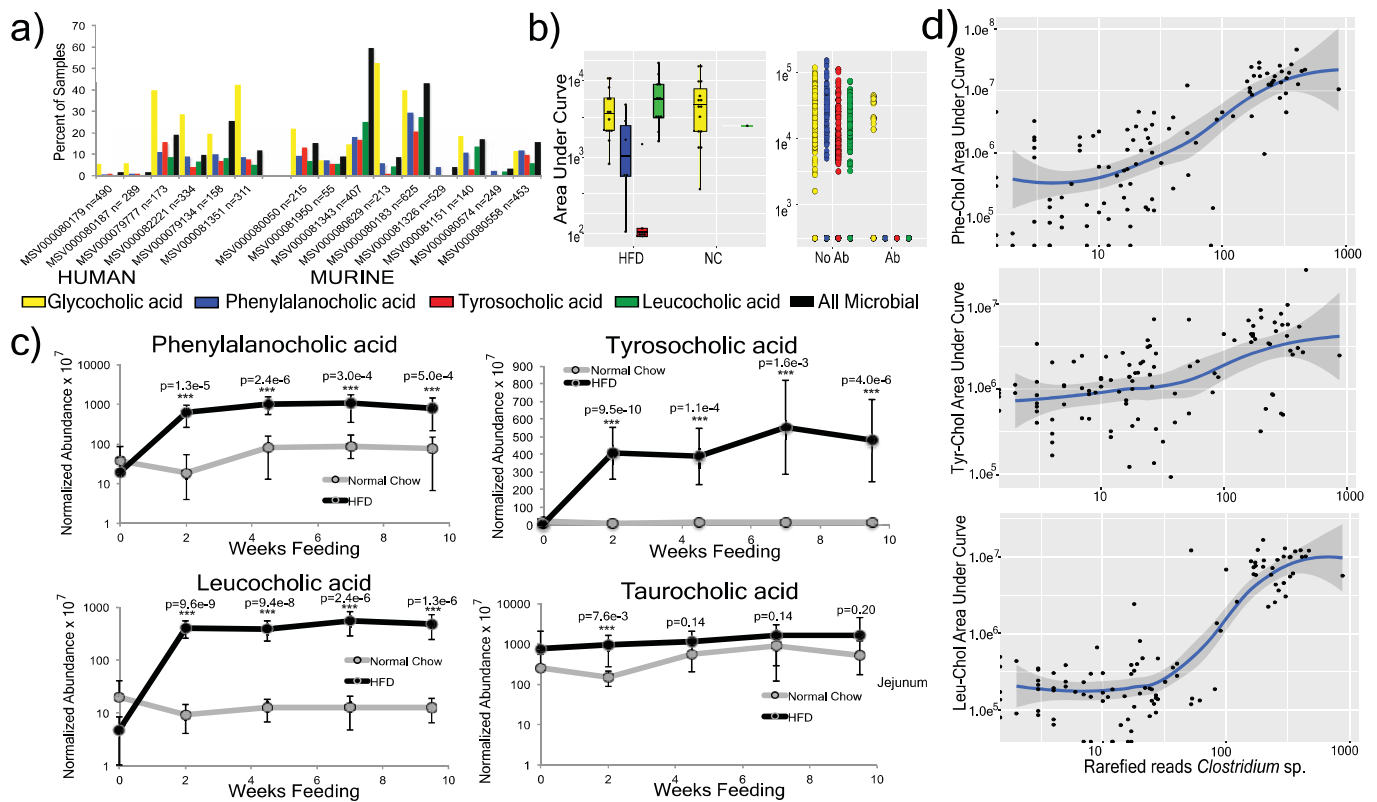


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Distribution and metabolism of newly identified

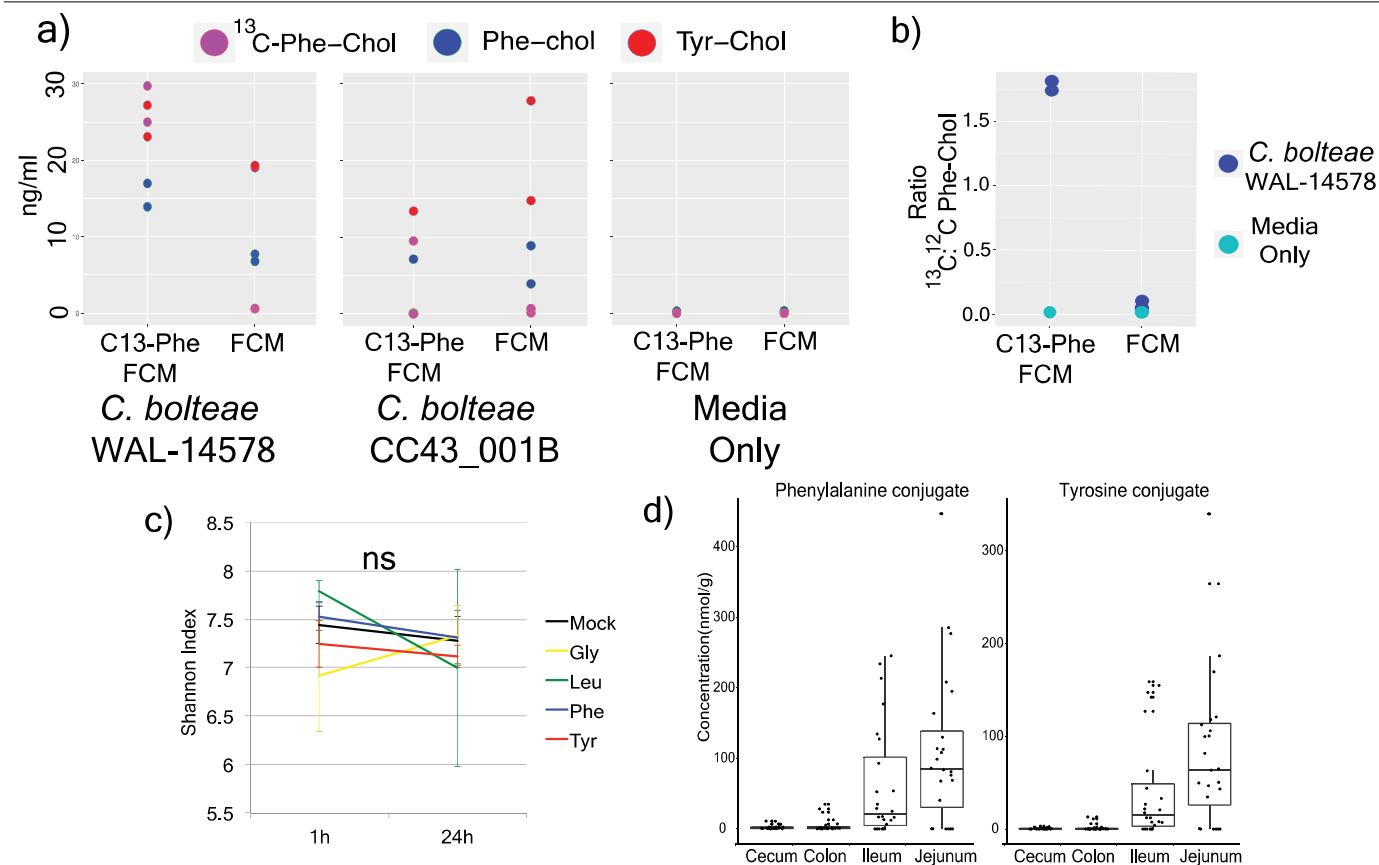
conjugated bile acids. **a**, Molecular network of MS/MS data from synthesized amino acid conjugated bile acids and the duodenum of SPF mice. LC-MS/MS data from synthetic standards were networked with mouse samples and spectral matching. Molecular networking is indicated by node colouring. Mirror plots show the alignment between the mouse and the synthetic standards. Nodes shaped as arrowheads had hits in the GNPS libraries, and node size is scaled to the spectral count. Tauro, taurocholic acid. These experiments were performed twice. **b**, Three-dimensional molecular cartography of the mean abundance of the newly discovered conjugates mapped onto a 3D-rendered model of the mouse gastrointestinal tract, as a heat map according to the colour scale. Organs are labelled as described in Fig. 1. **c**, Molecular network of conjugated bile acids from portal and peripheral blood of germ-free and SPF mice. Nodes are coloured by source as germ-free portal, germ-free portal and peripheral blood, SPF portal and peripheral blood, GF portal and peripheral blood and SPF peripheral blood, and all. Arrowhead

nodes represent known compounds in the GNPS spectral database; circular nodes represent unknown compounds. The annotations were obtained through spectral matches against reference libraries (level two or three¹⁶). **d**, Mean area-under-the-curve abundance and s.d. of bile acids of interest during incubation with an actively growing batch human faecal culture for 24 h ($n = 3$ independent incubations). **e**, Molecular network of newly identified conjugated bile acids after incubation in a human faecal batch culture experiment. Each node represents a unique tandem mass spectrum; arrowhead-shaped nodes indicate known spectra in the GNPS database. The nodes are coloured by their retention time according to the legend, and the mass shifts between nodes are mapped onto the edge representing the cosine connection between related spectra. The H_2 mass shift representing oxidation of the newly identified conjugates is shown. **f**, Mean ion intensity and s.d. of the oxidized forms of Phe-chol, Tyr-chol and Leu-chol through the 24-h batch faecal culture incubation ($n = 3$ independent incubations).



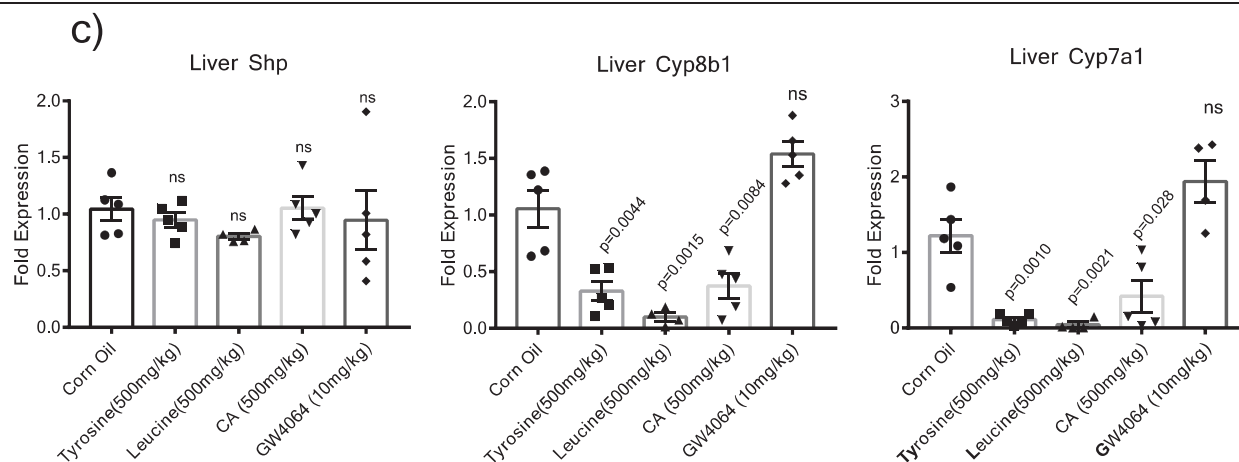
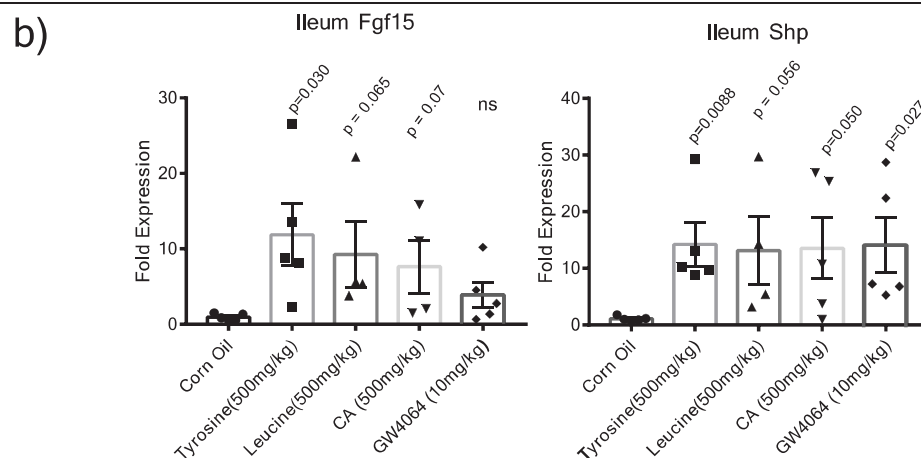
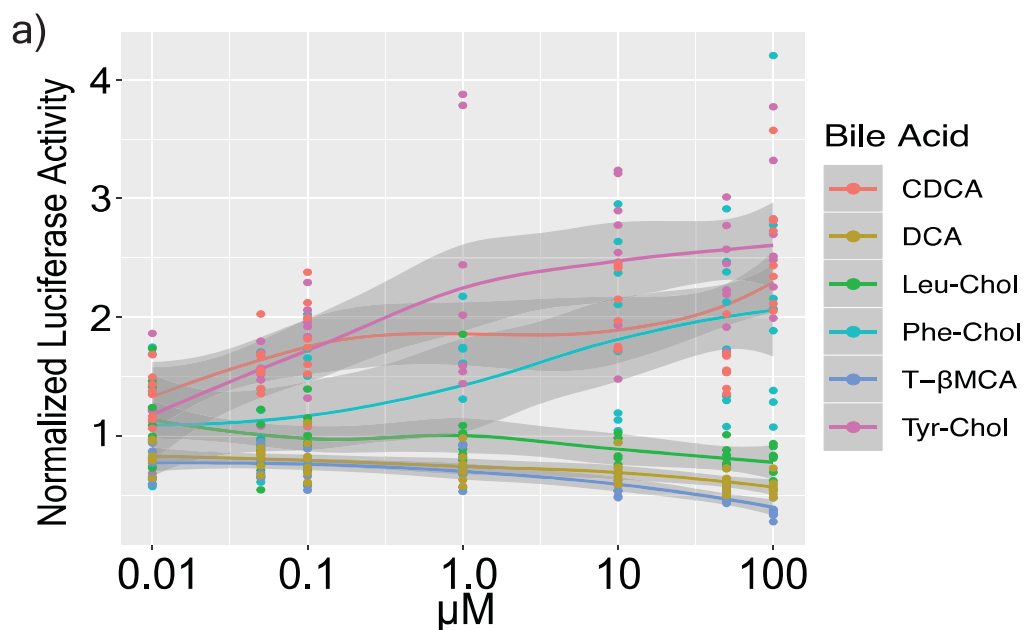
Extended Data Fig. 7 | MASST search results and associations of newly identified conjugated bile acids with high-fat diet. a, Proportion of samples in which Phe-chol, Tyr-chol and Leu-chol were found from a single-spectrum MASST search of publicly available data on GNPS. Massive dataset identifiers are shown for each dataset, are divided into mouse ('murine') or human gastrointestinal samples. **b,** Box plots of the newly identified conjugates in a previously published mouse study, in which mice were fed high-fat diet (HFD) ($n=14$ mice) or normal chow (NC) ($n=19$ mice) (Gly, $P=0.72$; Phe, $P=0.038$; Tyr, $P=0.083$; Leu $P=9.4 \times 10^{-5}$) and dot plot of mice treated with ($n=27$ mice) or without antibiotics (Ab) ($n=415$ mice)²⁹. Boxes represent the IQR, the line is the median and whiskers are $1.5 \times$ the IQR. Colour legend applies to both **a** and **b**. **c,** Mean normalized abundance of the three newly identified conjugated bile

acids compared to taurocholic acid in mice (*Apoe*-knockout on a C57BL/6J background) fed either a high-fat diet ($n=12$ mice) or normal chow ($n=12$ mice) for 10 weeks. Faecal samples were collected and extracted in 50:50 methanol:water and analysed with LC-MS/MS metabolomics, as described in the Supplementary Methods. The s.d. around the mean is shown, and significance between a high-fat diet and normal chow at each time point is tested with two-sided Student's *t*-test. *** $P < 0.001$. **d,** Correlations between rarefied reads of a deblurred read assigned to a *Clostridium* sp. from atherosclerosis-prone mice fed a high-fat diet over time ($n=12$ mice). The line of best fit is plotted using the lm method in the R statistical software; grey area around the line of best fit is the 95% confidence interval.



Extended Data Fig. 8 | Synthesis of newly identified conjugated bile acids by *Clostridium*. **a**, Dot plot of the measured production of Phe-chol and Tyr-chol using a targeted liquid chromatography–mass spectrometry method for two *C. bolteae* strains grown in faecal culture medium (FCM) with or without labelled Phe ($n = 2$ independent cultures). **b**, The mean ratio and s.e.m. of ^{13}C -Phe-chol: ^{12}C -Phe-chol from the same *C. bolteae* strains when grown with faecal culture medium with ^{13}C -labelled phenylalanine (bottom left)

($n = 2$ cultures). **c**, Mean and s.d. of the Shannon index of human faecal batch culture ($n = 3$ cultures) before and after 24-h growth exposed to conjugated bile acids or a mock control. NS, not significant by Mann–Whitney U -test. **d**, Box-and-whisker plots of concentration of Phe-chol and Tyr-chol in original samples from the gut of SPF mice. Boxes represent the IQR, the centre is the median and whiskers are 1.5× the IQR. $n = 4$ mice.



Extended Data Fig. 9 | Effect of newly identified bile acids on FXR. a, Mean normalized luciferase activity as a readout of human FXR stimulation when exposed to various conjugated and unconjugated bile acids, as a function of the compound dose. $n=8$ measurements, \pm s.e.m. DCA, deoxycholic acid; CDCA, chenodeoxycholic acid; T-βMCA, tauro-β-muricholic acid. **b,** Ileum

mean fold expression change compared to 36B4 control of various bile acids after gavage in mice. Error bars are s.e.m. **c,** Liver fold expression change compared to 36B4 control of various bile acids after gavage in mice. Significance was tested with two-tailed t -test compared to the mock corn-oil control. Error bars are s.e.m.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

'ili 3D-cartography modeling code is available here <https://github.com/MolecularCartography/ili>
Single spectrum search is available here https://gnps.ucsd.edu/ProteoSAFe/index.jsp?params=%7B%22workflow%22%3A%22SEARCH_SINGLE_SPECTRUM%22%7D
'ili 3D-cartography modeling code is available here <https://github.com/MolecularCartography/ili>
The MASST search custom code is available here: https://github.com/CCMS-UCSD/GNPS_Workflows/tree/master/search_single_spectrum
Invesalius Version 3 was used to process MRI data and 3D models were generated with the ili software available here: ili.embl.de

Data analysis

Invesalius software (<https://invesalius.github.io/>) was used for generation of mouse 3D-model from MRI dicoms.
'ili software was used for 3D-mapping <https://github.com/MolecularCartography/ili>
Invesalius Version 3 was used to process MRI data and 3D models were generated with the ili software available here: ili.embl.de

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All metabolomics data is available at GNPS (gnps.ucsd.edu) under the MassIVE ID numbers: MSV000079949 (GF and SPF mouse data). Additional MS datasets:

MSV000082480, MSV000082467, MSV000079134, MSV000082406, MSV000083032, MSV000083004, MSV000083446.

The sequencing data for the GF and SPF mouse study is available on the Qiita microbiome data analysis platform at Qiita.ucsd.edu under study ID 10801 and through the European Bioinformatics Institute accession number ERP109688.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No initial sample sizes were predetermined for the GF and SPF mice. We were searching for molecules present only in colonized mice our n=4 enabled a higher level of confidence that these molecules were not due to individual variation but in fact presence of a microbiome. Upon FXR agonism tests during gavage experiments initial data using n=4 was used to project required n for subsequent validation experiments. No sample sizes were calculated for the clinical experiments. All of these were searching of publicly available data for the new bile acids MS/MS spectra. We subsequently tested abundance of these compounds in the subjects datasets and if significance was found it was reported, thus, the sample sizes were sufficient.
Data exclusions	There are no data exclusions from the clinical experiments or any other aspect of the manuscript.
Replication	The multiple mice tested in each group enabled a verification that molecules found were consistent and reproducibly from the microbiome. Searching the public datasets also provided validation for the molecules association with guts of mammals. Other aspects of this study were broad scale mining of publicly available mass spectrometry data, thus, we did not attempt to reproduce the findings, however, links to the workflows and data are available such this can be easily replicated
Randomization	The datasets are randomized in the context of our molecular MASST searching. We have no prior knowledge of what any of the samples are and the searches report the results regardless of grouping of samples. But again, we don't believe mining public data has relevance in this context.
Blinding	The investigators were blinded to the study subjects. There was no prior knowledge of what studies the compounds were present in, and no groups of patients/subjects were known prior to the MS/MS searches.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Germ-free (GF) C57Bl/6J mice and Conventionally-colonized specific pathogen free (SPF) mice (C57Bl/6J)
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	The California Institute of Technology and University of California at San Diego provided guidance/oversight.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The human populations used were not relevant as covariates as they were only mined to identify the presence of a molecule in their sample sets. The cystic fibrosis patient population was a pediatric group age 0-18 years.
Recruitment	Patients were recruited during routine visits to UCSD clinics there was no bias towards selection of any particular group.
Ethics oversight	UCSD Institutional Review Board, University of Michigan, and Yale University Institutional Review Boards

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	<i>Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.</i>
Study protocol	<i>Note where the full trial protocol can be accessed OR if not available, explain why.</i>
Data collection	<i>Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.</i>
Outcomes	<i>Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.</i>

ILC2s amplify PD-1 blockade by activating tissue-specific cancer immunity

<https://doi.org/10.1038/s41586-020-2015-4>

Received: 26 March 2019

Accepted: 31 December 2019

Published online: 19 February 2020

 Check for updates

John Alec Moral^{1,2,3}, Joanne Leung^{1,2,3,17}, Luis A. Rojas^{1,2,3,17}, Jennifer Ruan^{1,2,3,17}, Julia Zhao^{1,2,3}, Zachary Sethna^{1,2,3}, Anita Ramnarain^{1,2,3}, Billel Gasmi⁴, Murali Gururajan⁵, David Redmond⁶, Gokce Askan⁷, Umesh Bhanot⁷, Ela Elyada^{8,9}, Youngkyu Park^{8,9}, David A. Tuveson^{8,9}, Mithat Gönen¹⁰, Steven D. Leach¹¹, Jedd D. Wolchok^{2,4,12,13,14,15}, Ronald P. DeMatteo¹⁶, Taha Merghoub^{2,4,13,14,15}✉ & Vinod P. Balachandran^{1,2,3,13}✉

Group 2 innate lymphoid cells (ILC2s) regulate inflammation and immunity in mammalian tissues^{1,2}. Although ILC2s are found in cancers of these tissues³, their roles in cancer immunity and immunotherapy are unclear. Here we show that ILC2s infiltrate pancreatic ductal adenocarcinomas (PDACs) to activate tissue-specific tumour immunity. Interleukin-33 (IL33) activates tumour ILC2s (TILC2s) and CD8⁺ T cells in orthotopic pancreatic tumours but not heterotopic skin tumours in mice to restrict pancreas-specific tumour growth. Resting and activated TILC2s express the inhibitory checkpoint receptor PD-1. Antibody-mediated PD-1 blockade relieves ILC2 cell-intrinsic PD-1 inhibition to expand TILC2s, augment anti-tumour immunity, and enhance tumour control, identifying activated TILC2s as targets of anti-PD-1 immunotherapy. Finally, both PD-1⁺ TILC2s and PD-1⁺ T cells are present in most human PDACs. Our results identify ILC2s as anti-cancer immune cells for PDAC immunotherapy. More broadly, ILC2s emerge as tissue-specific enhancers of cancer immunity that amplify the efficacy of anti-PD-1 immunotherapy. As ILC2s and T cells co-exist in human cancers and share stimulatory and inhibitory pathways, immunotherapeutic strategies to collectively target anti-cancer ILC2s and T cells may be broadly applicable.

ILC2s are innate antigen-independent lymphocytes that regulate immunity to pathogens and commensals in tissues. Although ILC2s have been detected in cancers, their role in tumour immunity is unclear.

TILC2s infiltrate pancreatic cancers

To investigate the role of ILC2s in cancer, we analysed tumour-infiltrating lymphocytes in unselected primary human PDACs. We found intra-tumoral cells that lacked immune cell lineage markers (lineage⁻) but expressed markers of ILCs (CD25 and CD127)² and ILC2s (IL33 receptor (ST2, also known as IL1RL1 or IL33R) and GATA3) (Fig. 1a, Extended Data Fig. 1a, Supplementary Table 1). These putative TILC2s were enriched in ‘hot’ tumours (enriched in CD8⁺ T cells) from rare long-term PDAC survivors⁴ when compared with ‘cold’ tumours from short-term survivors. In addition, higher TILC2 frequencies correlated with longer survival (Fig. 1b, Extended Data Fig. 1b). Higher bulk RNA expression of the ILC2-activating cytokine *IL33* in tumours, but not of any other

ILC-activating cytokine, was associated with longer survival (Fig. 1c, Extended Data Fig. 1c, Supplementary Table 2). Furthermore, expression of *IL33*, but not of other ILC-activating cytokines, correlated with higher intratumoral immune cytolytic activity (Fig. 1c, Extended Data Fig. 1c). Although these data assess RNA and not protein expression, they suggest that IL33 and TILC2s activate anti-tumour immunity in human PDAC.

We next looked for ILCs in tumours from mice in which PDAC development is driven by mutated *Kras* and *p53* (autochthonous KPC mice)⁵ and orthotopic mouse models of PDAC (PDAC mice^{6,7}). In both models, we detected TILC2s that were phenotypically similar to those in human PDACs and to mouse ILC2s^{2,8} (Fig. 1d, Extended Data Fig. 1d–f). The frequency of mouse ILC2s was increased in tumours, but not in adjacent organs (Fig. 1d, Extended Data Fig. 1g), consistent with their tissue residency⁹, and were depleted in *Rag2*^{-/-} mice^{10,11} by targeting the lymphocyte antigen CD90.2 (Fig. 1e, Extended Data Fig. 1h). Therefore, ILC2s are conserved cells that expand locally in mouse and human PDACs.

¹Hepatopancreatobiliary Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ²Parker Institute for Cancer Immunotherapy, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ³David M. Rubenstein Center for Pancreatic Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ⁴Swim Across America/Ludwig Collaborative Laboratory, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ⁵Immunology-Oncology Biology, Bristol Myers Squibb Company, Princeton, NJ, USA. ⁶Ansary Stem Cell Institute, Division of Regenerative Medicine, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. ⁷Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ⁸Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. ⁹Lustgarten Foundation Pancreatic Cancer Research Laboratory, Cold Spring Harbor, NY, USA. ¹⁰Department of Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹¹Dartmouth Norris Cotton Cancer Center, Lebanon, NH, USA. ¹²Melanoma and Immunotherapeutics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹³Weill Cornell Medical College, Cornell University, New York, NY, USA. ¹⁴Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹⁵Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹⁶Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ¹⁷These authors contributed equally: Joanne Leung, Luis A. Rojas, Jennifer Ruan.

✉e-mail: merghout@mskcc.org; balachav@mskcc.org

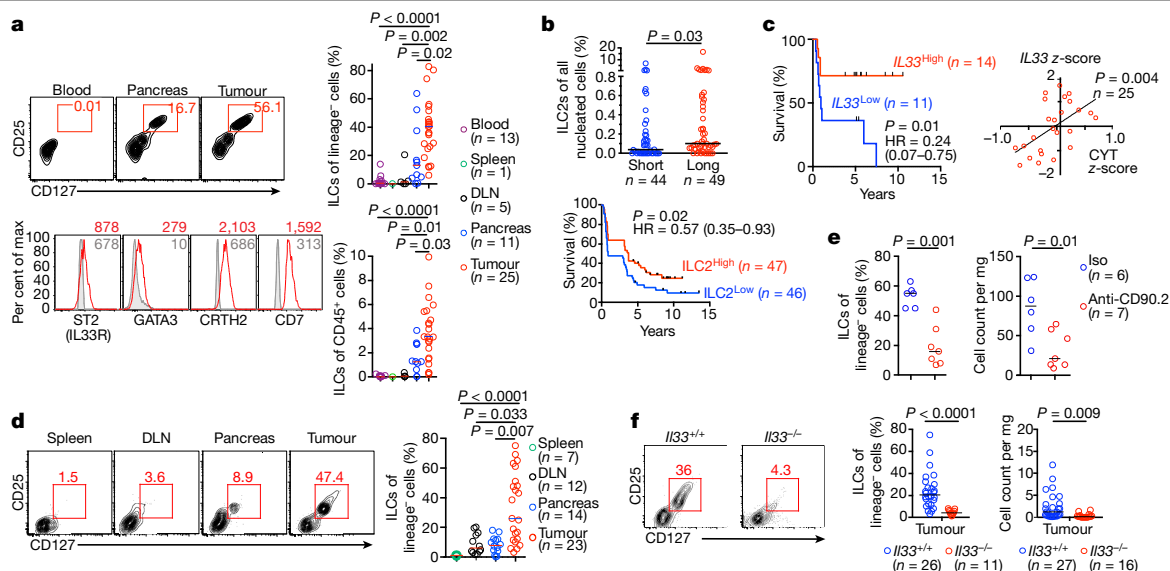


Fig. 1 | IL33-dependent TILC2s infiltrate human and mouse pancreatic cancer. **a**, Gating, frequency, and phenotype of ILCs from unselected patients with PDAC. Grey curves, isotype controls; numbers, mean fluorescence intensity. **b**, Frequency (top) and survival association (bottom) of ILC2s in microarrays of tumour samples from short- and long-term PDAC survivors. **c**, Bulk tumour *IL33* mRNA association with survival and tumour cytolytic index (CYT) in short- and long-term PDAC survivors. **d**, Gating and frequency of ILCs in mice with PDAC. **e**, Intratumoral ILC frequency and number in *Rag2*^{-/-} PDAC mice treated with anti-CD90.2 or isotype (Iso) antibodies. **f**, Gating, frequency,

and number of ILCs in *Il33*^{+/+} and *Il33*^{-/-} PDAC mice. High and low in **b**, **c** defined as higher or lower, respectively, than the median for the cohort. HR, hazard ratio. **d–f**, Data were collected 14 (**d**, **f**) and 10 (**e**) days after tumour implantation, pooled from two or more independent experiments with *n* ≥ 3 per group; each point indicates one mouse analysed separately. *n*, number of tumours from individual patients or mice. Horizontal bars show medians. *P* values determined by one-way ANOVA with Tukey's (**a**) and Kruskal–Wallis multiple comparison (**d**) post tests, two-tailed Mann–Whitney test (**b**, **e**, **f**), two-sided log-rank test (**b**, **c**, survival curves), and linear regression (**c**).

To identify the signals that expand TILC2s, we found that IL33 was the most highly expressed ILC-activating cytokine in tumours in both PDAC and KPC mice¹² (Extended Data Fig. 2a). IL33 was heterogeneously expressed in both human and mouse PDACs (Extended Data Fig. 2b, c) and maximally expressed in intratumoral myeloid cells^{13,14} (Extended Data Fig. 2d, e). To understand the role of IL33 and TILC2s in PDAC immunity, we studied TILC2 dependency on IL33 in *Il33*^{High} PDAC mice, to model *Il33*^{High}, ILC2-enriched hot tumours in long-term human PDAC survivors. The expansion and function of TILC2s were IL33-dependent, as *Il33*^{-/-} PDAC mice had reduced TILC2 number, frequency (Fig. 1f, Extended Data Fig. 2f) and cytokine production (Extended Data Fig. 2g) when compared to *Il33*^{+/+} PDAC mice. Recombinant IL33 (rIL33) expanded ILCs in ILC-proficient *Rag2*^{-/-} PDAC mice, but not in ILC-deficient mice that lack both *Rag2* and the gene encoding the gamma subunit of the IL-2 receptor (*Il2rg*, also known as *yc*) (*Rag2*^{-/-}*yc*^{-/-} PDAC mice) (Extended Data Fig. 2h, i). Collectively, these data show that IL33 expands TILC2s in PDACs.

TILC2s boost tumour immunity in tissues

As ILC2s have tissue-specific phenotypes¹⁵, we hypothesized that TILC2s have tissue-specific effects on PDAC immunity. To test this, we compared the effects of IL33 deficiency on pancreatic and skin tumour growth (pancreatic TILC2s express ST2 whereas skin TILC2s do not^{15,16}; Extended Data Fig. 2j). Compared with *Il33*^{+/+} mice, *Il33*^{-/-} mice with orthotopic PDAC had larger tumours, accelerated tumour growth, and worse survival (Fig. 2a). By contrast, mice with subcutaneous PDACs showed no IL33-dependent phenotype (Fig. 2b, Extended Data Fig. 2k). Although these mice were fully backcrossed onto identical genetic backgrounds, we confirmed that the differences were not due to potential minor genetic mismatches by observing larger tumours in *Il33*^{-/-} mice compared to *Il33*^{+/+} littermates (Extended Data Fig. 2l). These anti-tumour effects depended on IL33 produced by host haematopoietic cells, as chimaeric mice transplanted with *Il33*^{-/-} bone marrow had larger tumours than did control mice (Extended Data Fig. 2m–o). RNA

sequencing (RNA-seq) of purified CD45⁺ intratumoral immune cells from *Il33*^{+/+} and *Il33*^{-/-} mice with orthotopic PDACs showed that PDAC immune cells from *Il33*^{-/-} mice had diminished transcriptional signatures of T cell activation and MHC class I antigen processing (Extended Data Fig. 3a), suggesting that *Il33*^{-/-} PDAC mice might have a defect in T cell priming. Consistently, *Il33*^{-/-} mice with orthotopic but not subcutaneous PDACs had lower frequencies of global and activated tumour-infiltrating CD8⁺ T cells and reduced central memory CD8⁺ T cells (T_{CM}) in draining lymph nodes (DLN) but not distant lymph nodes, with no consistent changes in other immune cell frequencies (Fig. 2c, Extended Data Fig. 3b–e). Depletion of all T cells prevented the increase in tumour size in *Il33*^{-/-} compared to *Il33*^{+/+} mice (Fig. 2d), and rIL33-treated *Rag2*^{-/-} PDAC mice showed no differences in tumour weight compared with untreated mice (Extended Data Fig. 4a), confirming that the anti-tumour effects of IL33 were mediated by T cells. Orthotopic tumours from *Il33*^{-/-} and *Il33*^{+/+} PDAC mice also had similar histology and collagen and fibroblast content (Extended Data Fig. 4b–d), and rIL33 had no effects on tumour cells in vitro (Extended Data Fig. 4e–g), showing that IL33 had no direct effects on tumour or stromal cells. Together, these data show that IL33 activates tissue-specific cancer immunity by potentially activating TILC2s to prime CD8⁺ T cells.

We next tested whether the tissue-specific effects of IL33 depended on CD8⁺ T cells by contrasting the rejection phenotypes of KPC cells expressing the CD8⁺ T cell rejection antigen ovalbumin (KPC-OVA cells) at different tissue sites. Notably, 70% of *Il33*^{+/+} mice rejected orthotopic KPC-OVA tumours, whereas 0% of *Il33*^{-/-} mice did. By contrast, 100% of *Il33*^{+/+} and *Il33*^{-/-} mice rejected subcutaneous KPC-OVA tumours (Fig. 2e). To assess whether this phenotype resulted from ILC2 deficiency and ineffective CD8⁺ T cell priming, we acutely depleted ILC2s and examined antigen-specific CD8⁺ T cells in DLNs in the iCOS-T mouse, in which diphtheria toxin depletes ILC2s while sparing ICOS⁺CD4⁺ T cells¹⁷ (Fig. 2f, Extended Data Fig. 5a). ILC2 depletion recapitulated the *Il33*^{-/-} phenotype; mice with orthotopic KPC-OVA tumours showed a lower rate of tumour rejection and larger tumour size, whereas those with subcutaneous tumours showed no differences (Fig. 2f), with an

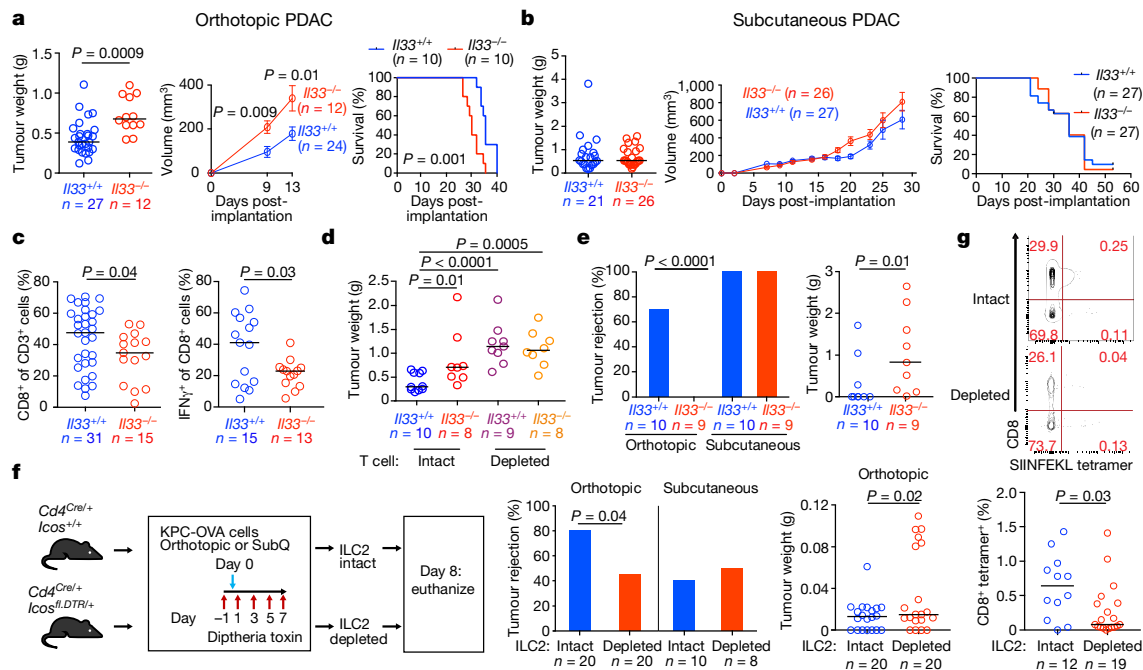


Fig. 2 | The IL33-ILC2 axis activates tissue-specific cancer immunity.

a, b, Tumour weight, tumour volume, and survival of *Il33^{+/+}* and *Il33^{-/-}* mice with orthotopic (**a**) or subcutaneous (**b**) PDACs. **c**, Frequency of all (left) and IFN- γ producing (right) CD8⁺ T cells in orthotopic *Il33^{+/+}* and *Il33^{-/-}* PDAC tumours. **d**, Tumour weight in T cell-depleted *Il33^{+/+}* and *Il33^{-/-}* mice with orthotopic PDACs. **e**, Frequency of tumour rejection (orthotopic and subcutaneous) and tumour weight (orthotopic) in *Il33^{+/+}* and *Il33^{-/-}* mice with KPC-OVA PDACs. **f**, Experimental design (left), frequency of tumour rejection (middle), and tumour weight (right) of KPC-OVA PDAC tumours in iCOS-T mice with intact or

depleted ILC2s. **g**, Gating (top) and frequency (bottom) of OVA-specific CD8⁺ T cells in DLNs of iCOS-T mice with intact or depleted ILC2s in orthotopic KPC-OVA PDACs. Data were collected 14 days (**a, c, d**), 28 days (**b**), 42 days (**e**), and 8 days (**f, g**) after implantation. Mean \pm s.e.m; horizontal bars show median. Data pooled from two or more independent experiments with *n* \geq 4 per group; *n* and data points denote individual mice analysed separately. *P* values determined by two-tailed Mann-Whitney test (**a–g**), two-sided log-rank test (**a, b**, survival curves), two-way ANOVA with Sidak's multiple comparison test (**a, b**, tumour curves), and χ^2 test (**e, f**, per cent rejection).

anticipated varied phenotype compared to *Il33^{-/-}* mice due to differences in time of rejection assessment and depletion efficacy. Tetramer analysis in mice with orthotopic KPC-OVA tumours depleted of ILC2s revealed a reduced frequency of OVA-specific CD8⁺ T cells in DLNs and spleens, and a reduced frequency of CD8⁺ T_{CM} cells in DLNs (as seen in *Il33^{-/-}* mice) (Fig. 2g, Extended Data Fig. 5b, c). Therefore, ILC2 deficiency partially phenocopied *Il33* deficiency. Although direct effects of *Il33* on CD8⁺ T cells cannot be ruled out, we found no ST2 expression on intratumoral CD8⁺ T cells (Extended Data Fig. 5d). To summarize, these loss-of-function experiments suggest that the IL33-TILC2 axis primes tissue-specific CD8⁺ T cell PDAC immunity.

Next, we investigated whether rIL33 treatment had similar tissue-specific anti-tumour effects. Treatment with rIL33 prevented tumour establishment in mice with orthotopic PDACs and prolonged survival, but had no effects on mice with subcutaneous PDACs, resulting in progressive tumour growth and ulceration requiring euthanasia (Fig. 3a). rIL33 had similar tissue-specific anti-tumour effects in KPC-OVA PDAC mice (Extended Data Fig. 6a). Similarly, rIL18, a cytokine that preferentially activates IL18R⁺ skin ILC2s¹⁵, restricted the growth of subcutaneous PDACs infiltrated by IL18R⁺ ILCs, but not of orthotopic PDACs, which lack IL18R⁺ ILCs (Fig. 3b, Extended Data Fig. 6b). rIL33 selectively expanded ILC2s in DLNs and tumours (Fig. 3c) but not in spleens of mice with orthotopic PDACs or in any organs in mice with subcutaneous PDACs (Extended Data Fig. 6c, d). ILC2 expansion was accompanied by enhanced intratumoral CD8⁺ T cell cytokine capacity and PD-1 upregulation (Extended Data Fig. 6e), with no consistent changes in other intratumoral immune cells (Extended Data Fig. 6f), although potential modulation of their function cannot be ruled out. Consistent with indirect priming of anti-tumour CD8⁺ T cells by ILC2s, rIL33 treatment doubled the number of intratumoral CD103⁺ dendritic cells (DCs) (Fig. 3d, Extended Data Fig. 6g), which prime and recruit

CD8⁺ T cells into PDACs⁷. To determine whether the effects of rIL33 depended on ILC2s, we administered rIL33 to PDAC-bearing *Rora^{fl/fl} Il7^{Cre/+}* mice, which are constitutively deficient in ILC2s¹⁷. ILC2 deficiency (Extended Data Fig. 6h) abrogated the efficacy of rIL33 (Fig. 3e) and attenuated increases in CD103⁺ DCs in tumours (Fig. 3f). rIL33 also had no anti-tumour effects (Fig. 3g) and failed to induce PD-1 expression in intratumoral CD8⁺ T cells (Extended Data Fig. 6i) in CD103⁺ DC-deficient *Batf3^{-/-}* mice, showing that CD103⁺ DCs are essential for rIL33-mediated tumour control. To test whether TILC2s produced chemokines to recruit DCs into tumours, we used single-cell RNA-seq (scRNA-seq; Extended Data Fig. 7a–c, Supplementary Table 3) and found that rIL33-activated TILC2s retained markers of ILC2 identity but exhibited distinct transcriptional profiles (Extended Data Fig. 8a–e), and selectively expressed *Ccl5* (Extended Data Fig. 8f). CCL5 recruits CD103⁺ DCs into tumours¹⁸ and induced efficient DC migration in vitro (Fig. 3h). Together, these data suggest that rIL33 expands TILC2s to produce CCL5, potentially recruit CD103⁺ DCs into tumours, and activate CD8⁺ T cells to induce therapeutic tumour immunity.

PD-1 blockade activates TILC2s

As stimulating ILC2s with rIL33 had anti-tumour effects, we searched for strategies to further activate ILC2s. Recent data have shown that, like T cells, ILC2s regulate their activity through coinhibitory immune checkpoint pathways. Specifically, the immune checkpoint PD-1 regulates mouse ILC2 development and marks effector ILCs¹⁹, and when PD-1 is genetically deficient or inhibited with a blocking antibody (anti-PD-1), IL33-activated ILC2s show increased expansion and effector function in mice and humans²⁰. PD-1⁺ ILC2s are also found in human tumours³. However, inhibiting immune checkpoints on ILC2s for cancer therapy has been unexplored.

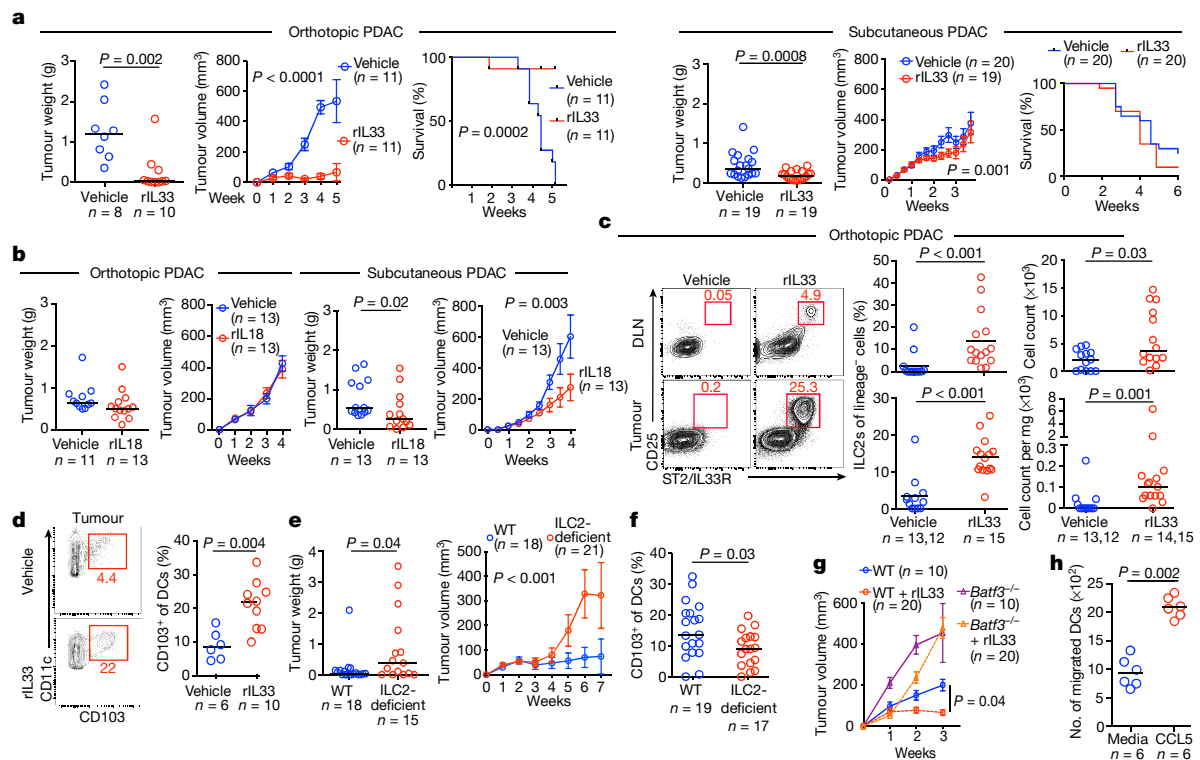


Fig. 3 | ILC2s stimulate tissue-specific cancer immunity by recruiting intratumoral dendritic cells. **a**, Tumour weight, tumour volume, and survival in orthotopic and subcutaneous PDAC mice treated with vehicle or rIL33. **b**, Tumour weight and volume in mice with orthotopic or subcutaneous PDACs treated with vehicle or rIL18. **c**, Gating, frequency, and number of ILC2s in rIL33-treated mice with orthotopic PDACs (DLN vehicle, $n = 13$, top; tumour vehicle, $n = 12$, bottom; DLN rIL33, $n = 14$, top; tumour rIL33, $n = 15$, bottom). **d**, Gating and frequency of CD103⁺ DCs in tumours from rIL33-treated mice with orthotopic PDACs. **e**, **f**, Tumour weight and volume (**e**) and frequency of CD103⁺

DCs (**f**) in tumours from rIL33-treated wild-type (WT) and ILC2-deficient mice with orthotopic PDACs. **g**, Tumour volume in rIL33-treated WT and CD103⁺ DC-deficient *Batf3*^{-/-} mice with orthotopic PDACs. **h**, Migration of purified DCs towards CCL5. Data were collected 5 (**c**, **d**) or 7 (**e**, **f**) weeks after tumour implantation. Median \pm s.e.m; horizontal bars show median. Data were pooled from at least two independent experiments, with $n \geq 3$ per group; n and data points denote individual mice analysed separately or individual replicates (**h**). *P* values determined by two-sided log-rank test (**a**, survival curves), two-way ANOVA (**a**, **b**, **e**, **g**, tumour volume), and two-tailed Mann–Whitney test (**a**–**f**, **h**).

Using scRNA-seq (Extended Data Fig. 7a–c), we found that PD-1 was the only detectable coinhibitory molecule expressed at baseline by TILC2s (Extended Data Fig. 9a). Treatment with rIL33 upregulated PD-1 on a fraction of TILC2s but not on DLN ILC2s (Extended Data Fig. 9b), suggesting that PD-1 may functionally restrain activated TILC2s. We therefore tested whether combining rIL33 with anti-PD-1 treatment could cooperatively activate TILC2s to enhance anti-tumour efficacy. Consistent with the expression of PD-1 only on rIL33-activated TILC2s, anti-PD-1 alone induced a partial anti-tumour response (Fig. 4a, as previously reported in PDACs⁷) but did not appreciably alter TILC2 frequencies (Fig. 4b, Extended Data Fig. 9c). A combination of rIL33 and anti-PD-1 maximally expanded ILC2s in tumours and DLNs (Fig. 4b) and enhanced tumour control compared to anti-PD-1 alone (Fig. 4a). To investigate whether anti-PD-1 activated ILC2s by cell-intrinsic PD-1 blockade, we compared the single-cell transcriptional profiles of TILC2s and DLN ILC2s following *in vivo* treatment. Whereas TILC2s retained the transcriptional and cellular identities of ILC2s irrespective of treatment (Extended Data Fig. 9d), rIL33 and anti-PD-1 induced a unique transcriptional phenotype in TILC2s in PDAC mice compared to all other treatment conditions and tissue sites (Fig. 4c), and increased expression of ILC2 markers, canonical (amphiregulin (*Areg*)¹⁵ and non-canonical (*Cxcl2*)²¹ ILC2 effector molecules, cellular activation machinery (*Junb*, *Fosl2*, and *Ybx1*), and coinhibitory immune checkpoints (Extended Data Fig. 9e–i). Finally, the anti-tumour effects of dual therapy were abrogated in ILC2-deficient mice (Fig. 4d), which shows that ILC2s are necessary for the efficacy of dual anti-PD-1 and rIL33 therapy. These results suggest that anti-PD-1 amplifies activated TILC2s by possibly inhibiting the PD-1 pathway on ILC2s, and not on T cells.

PD-1 inhibits cell-intrinsic TILC2 function

To investigate whether interrupting the PD-1 pathway on activated TILC2s contributed to the anti-tumour effects of dual therapy, we transferred sort-purified rIL33-activated PD-1 proficient (wild-type) or PD-1 deficient (*Pdcd1*^{-/-}) TILC2s into tumour-bearing ILC2-deficient mice (Fig. 4e, Extended Data Fig. 10a). Transfer of wild-type TILC2s had no anti-tumour efficacy in established tumours, but transfer of *Pdcd1*^{-/-} TILC2s restricted tumour growth, indicating that interrupting PD-1 signalling on TILC2s can enhance tumour control (Fig. 4e). We next tested whether rIL33-activated PD-1⁺ TILC2s could directly amplify the efficacy of anti-PD-1 therapy in established tumours. We transferred sort-purified rIL33-activated congenic CD45.1⁺ TILC2s into CD45.2⁺ ILC2-deficient mice with established tumours, and treated the mice with anti-PD-1 after the transfer (Fig. 4f). Transferred TILC2s were more than 97% PD-1⁺ (Extended Data Fig. 10b), accumulated in the tumours and DLNs but not spleens of anti-PD-1 treated recipient mice, and persisted for up to 9 weeks post-transfer (Fig. 4g). Transfer of PD-1⁺ TILC2s augmented the efficacy of anti-PD-1 treatment, restricted tumour growth (Fig. 4f), and increased T cell frequencies in the tumours and DLNs, but not spleens, of recipient mice (Fig. 4h). These data show that blocking PD-1 signalling on rIL33-activated TILC2s directly amplified the anti-tumour efficacy of anti-PD-1 treatment.

To examine the efficacy of rIL33 and anti-PD-1 treatment in IL33^{low}, anti-PD-1-resistant tumours, we selected an aggressive cold PDAC tumour model (KPC S2 mice) that generates IL33^{low} tumours (Extended Data Fig. 2b), has 50% fewer CD8⁺ T cells than IL33^{high} tumours, and has a median survival of only 2 weeks (Extended Data Fig. 10c), to mimic

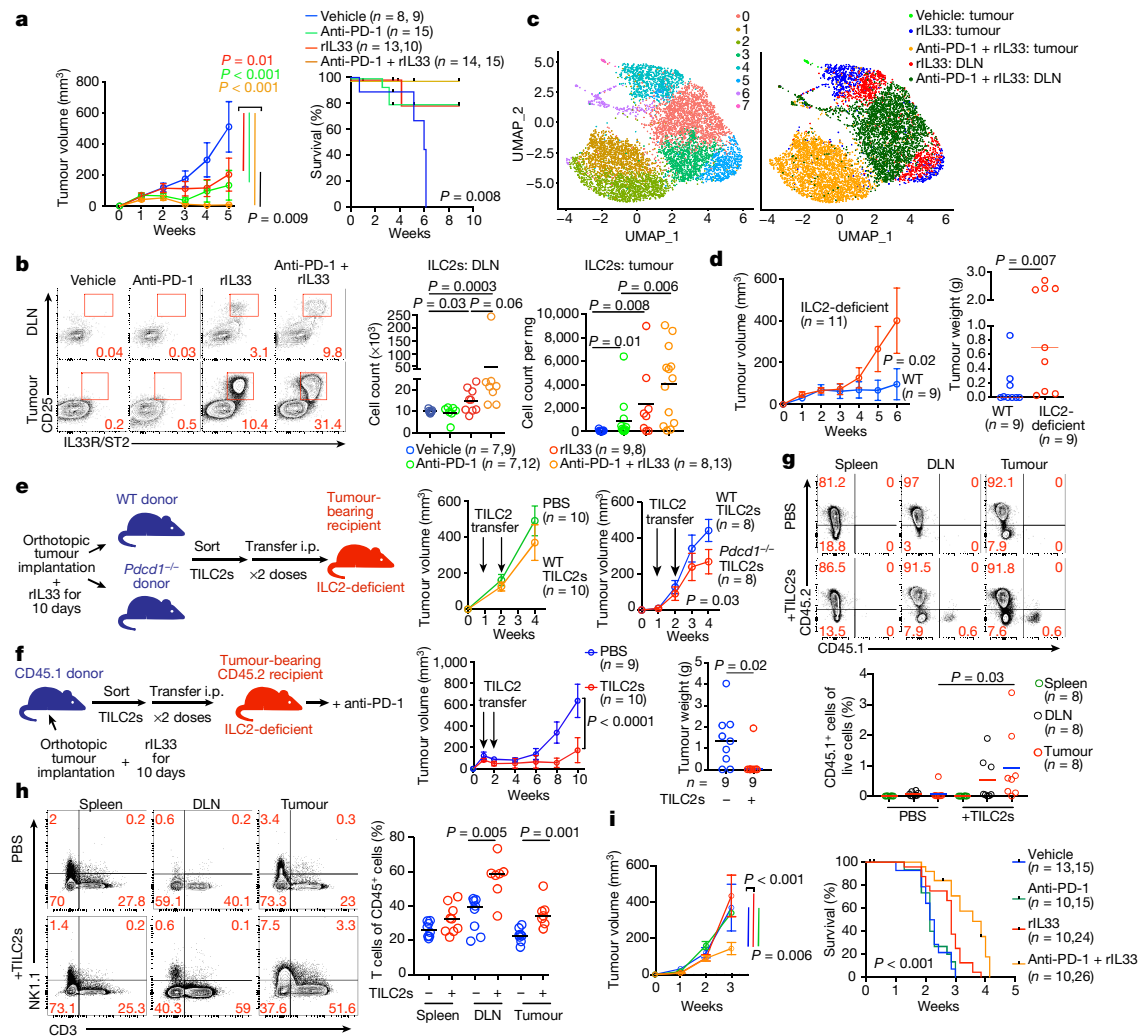


Fig. 4 | PD-1 blockade activates TILC2s. **a**, **b**, Tumour volume (vehicle, $n = 8$; rIL33, $n = 13$; anti-PD-1 + rIL33, $n = 14$) and survival (vehicle, $n = 9$; rIL33, $n = 10$; anti-PD-1 + rIL33, $n = 15$) (**a**) and gating, frequency, and number of ILC2s (**b**) in treated PDAC mice. **c**, scRNA-seq ($n = 7,022$ single ILC2s) in treated PDAC mice in a nonlinear representation of the top 15 principal components; cells are coloured by cluster (left) or treatment and tissue (right). **d**, Tumour volume in wild-type (WT) and ILC2-deficient PDAC mice treated with anti-PD-1 and rIL33. **e**, TILC2s were sort-purified from rIL33-treated WT or *Pdcd1*^{-/-} PDAC mice and transferred into ILC2-deficient PDAC mice, and tumour volumes were measured. **f**, **h**, TILC2s were sort-purified from rIL33-treated PDAC CD45.1 donor mice and transferred into ILC2-deficient CD45.2 PDAC recipient mice, which were then treated with anti-PD-1. **f**, Experimental design, tumour volume and tumour weight; **g**, frequency of CD45.1 and CD45.2 cells; **h**, frequency of T cells in recipient mice 9 weeks after cell transfer (TILC2⁺: all groups, $n = 8$;

TILC2⁺: spleen, $n = 9$; DLN, $n = 7$; tumour, $n = 7$). Frequencies in **g** represent percentage of live donor- or recipient-derived immune cells. **i**, Tumour volume (vehicle, $n = 13$; other groups, $n = 10$) and survival (vehicle and anti-PD-1, $n = 15$; rIL33, $n = 24$; rIL33 + anti-PD-1, $n = 26$) of treated PDAC mice (KPC S2 cells). ILC2 deficient, *Rora*^{fl/fl} *Il7*^{Cre}. Data were collected 5 weeks (**b**), 10 days (**c**), or 6 weeks (**d**) after orthotopic tumour cell implantation. Median \pm s.e.m; horizontal bars show median. Data pooled from at least two independent experiments with $n \geq 3$ per group; n and data points denote individual mice analysed separately. Data for scRNA-seq represent pooled purified single cells from biological replicates (vehicle $n = 10$, rIL33 $n = 5$, anti-PD-1 + rIL33 $n = 5$). P values determined by two-way ANOVA with Tukey's multiple comparison post test (**a**, **d**–**f**, **i**, tumour volume), two-tailed Mann–Whitney test (**b**, **d**, **f**, **g**, **h**), and two-sided log-rank test (**a**, **i**, survival curves).

the immunological and survival features of patients with IL33^{Low} PDAC tumours who show short-term survival. Although KPC S2 PDAC mice do not exhibit the sequential steps of PDAC tumorigenesis from pre-invasive neoplasias to invasive PDAC that are seen in spontaneous KPC mice, they recapitulated the anti-PD-1 resistance seen in spontaneous KPC mice and human PDACs (Fig. 4i). Combination treatment with rIL33 and anti-PD-1 reduced tumour volume by over 50% in these mice, with a nearly 50% improvement in survival (Fig. 4i). Finally, to assess the potential to use dual rIL33 and anti-PD-1 therapy to treat patients with PDAC, we investigated the co-occurrence of PD-1⁺ TILC2s and PD-1⁺ T cells in human PDACs. Nearly 60% of human PDACs had low frequencies of PD-1⁺ TILC2s and PD-1⁺ T cells, with a significant correlation between the two cell types (Extended Data Fig. 10d), which suggests that they frequently co-occur in human PDAC. In addition, *IL33* mRNA correlated

substantially with *PDCDI* mRNA, which encodes PD-1 (Extended Data Fig. 10e). PD-1 expression has been associated with longer survival²², which suggests that the IL33–PD-1 axis may positively impact survival in individuals with PDAC. In summary, rIL33-activated ILC2s can amplify responses to anti-PD-1 in both tumours that are partially sensitive to PD-1 and those that are PD-1-resistant.

Discussion

Our results suggest that ILC2s can be activated as a broader strategy to prime CD8⁺ T cells in cancers (Extended Data Fig. 10f). However, given the tissue-specific phenotypes of ILC2s, more work is required to determine whether activating them will have similar effects in cancers arising in different tissues. Given the divergent effects of ILC2s on

tumour immunity in different tissues, our findings also underscore the need for pre-clinical cancer studies to be performed in native organs to accurately reflect the local immune environments.

Immune checkpoints modulate ILC2 function, but the ability to harness ILC2s with immune checkpoint blockade for cancer therapy has remained unclear. We have shown that blocking PD-1 on activated ILC2s promotes anti-tumour effects, suggesting that ILC2s may partially contribute to the efficacy of PD-1 pathway blockade in human cancers. More broadly, this highlights that differential responses to immune checkpoint blockade may depend on tissue-specific factors. Refining strategies to identify ILC2s in human cancers will clarify their prognostic and predictive potential.

As activated ILC2s (Extended Data Fig. 10g) and T cells share several immune modulatory molecules and co-exist in human cancers, a broader array of checkpoints could be co-targeted on ILC2s and T cells in tumours. Further investigations to collectively target ILC2s and T cells for cancer immunotherapy are therefore warranted.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2015-4>.

1. Vivier, E., van de Pavert, S. A., Cooper, M. D. & Belz, G. T. The evolution of innate lymphoid cells. *Nat. Immunol.* **21**, 790–794 (2016).
2. Vivier, E. et al. Innate lymphoid cells: 10 years on. *Cell* **174**, 1054–1066 (2018).
3. Salimi, M. et al. Activated innate lymphoid cell populations accumulate in human tumour tissues. *BMC Cancer* **18**, 341 (2018).
4. Balachandran, V. P. et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
5. Hingorani, S. R. et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).

6. Pylayeva-Gupta, Y., Lee, K. E., Hajdu, C. H., Miller, G. & Bar-Sagi, D. Oncogenic Kras-induced GM-CSF production promotes the development of pancreatic neoplasia. *Cancer Cell* **21**, 836–847 (2012).
7. Li, J. et al. Tumor cell-intrinsic factors underlie heterogeneity of immune cell infiltration and response to immunotherapy. *Immunity* **49**, 178–193.e7 (2018).
8. Brestoff, J. R. et al. Group 2 innate lymphoid cells promote beiging of white adipose tissue and limit obesity. *Nature* **519**, 242–246 (2015).
9. Gasteiger, G., Fan, X., Dikiy, S., Lee, S. Y. & Rudensky, A. Y. Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. *Science* **350**, 981–985 (2015).
10. Kirchberger, S. et al. Innate lymphoid cells sustain colon cancer through production of interleukin-22 in a mouse model. *J. Exp. Med.* **210**, 917–931 (2013).
11. Monticelli, L. A. et al. Arginase 1 is an innate lymphoid-cell-intrinsic metabolic checkpoint controlling type 2 inflammation. *Nat. Immunol.* **17**, 656–665 (2016).
12. Neesse, A. et al. CTGF antagonism with mAb FG-3019 enhances chemotherapy response without increasing drug delivery in murine ductal pancreas cancer. *Proc. Natl Acad. Sci. USA* **110**, 12325–12330 (2013).
13. Hardman, C. S., Panova, V. & McKenzie, A. N. J. IL-33 citrine reporter mice reveal the temporal and spatial expression of IL-33 during allergic lung inflammation. *Eur. J. Immunol.* **43**, 488–498 (2013).
14. Talabot-Ayer, D. et al. The mouse interleukin (IL)33 gene is expressed in a cell type- and stimulus-dependent manner from two alternative promoters. *J. Leukoc. Biol.* **91**, 119–125 (2012).
15. Ricardo-Gonzalez, R. R. et al. Tissue signals imprint ILC2 identity with anticipatory function. *Nat. Immunol.* **19**, 1093–1099 (2018).
16. Dalmas, E. et al. Interleukin-33-activated islet-resident innate lymphoid cells promote insulin secretion through myeloid cell retinoic acid production. *Immunity* **47**, 928–942.e7 (2017).
17. Oliphant, C. J. et al. MHCII-mediated dialog between group 2 innate lymphoid cells and CD4⁺ T cells potentiates type 2 immunity and promotes parasitic helminth expulsion. *Immunity* **41**, 283–295 (2014).
18. Böttcher, J. P. et al. NK cells stimulate recruitment of cDC1 into the tumor microenvironment promoting cancer immune control. *Cell* **172**, 1022–1037.e14 (2018).
19. Yu, Y. et al. Single-cell RNA-seq identifies a PD-1^{hi} ILC progenitor and defines its development pathway. *Nature* **539**, 102–106 (2016).
20. Taylor, S. et al. PD-1 regulates KLRG1⁺ group 2 innate lymphoid cells. *J. Exp. Med.* **214**, 1663–1678 (2017).
21. Kim, J. et al. Intratumorally establishing type 2 innate lymphoid cells blocks tumor growth. *J. Immunol.* **196**, 2410–2423 (2016).
22. Diana, A. et al. Prognostic value, localization and correlation of PD-1/PD-L1, CD8 and FOXP3 with the desmoplastic stroma in pancreatic ductal adenocarcinoma. *Oncotarget* **7**, 40992–41004 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Mice

C57BL/6 (wild-type, WT, CD45.2), C57BL/6 CD45.1, *Rag2*^{-/-}, *Rag2*^{-/-}γc^{-/-}, *Batf3*^{-/-}, and *Pdcd1*^{-/-} mice were purchased from Jackson Labs. *Il33*^{-/-} and *Il33*^{Cre/+} mice were a gift from M. J. Rosen. *Cd4*^{Cre/+} *Icos*^{fl-DTR/+} and *Rora*^{fl/fl} *Il7*^{Cre/+} mice were a gift from A. N. J. McKenzie and have been previously described^{17,23}. For all experiments, 6–12-week-old mice were matched by age and sex and randomly assigned to specific treatment groups, with at least two independent experiments performed throughout. *Pdx1*^{Cre} *LSL-Kras*^{G12D/+} *LSL-Trp53*^{R172H/+} (KPC) mice have been previously described⁵. Sample sizes for experiments were determined without formal power calculations. Animals were bred and maintained in a specific pathogen-free animal facility, and all experiments were conducted in accordance with an Institutional Animal Care and Use Committee (IACUC) approved protocol at Memorial Sloan Kettering Cancer Center (MSKCC) and in compliance with all relevant ethical regulations.

Cell lines and animal procedures

All tumour cell lines were derived from KPC mice. KPC 4662 cells from *Pdx1*^{Cre} *LSL-Kras*^{G12D/+} *LSL-Trp53*^{R172H/+} mice (a gift from R. H. Vonderheide) were transfected with GFP and used for all experiments unless indicated otherwise. KPC 8-1, 18-3, and 52 cells derived from *Ptf1a*^{Cre} *LSL-Kras*^{G12D/+} *LSL-Trp53*^{R172H/+} mice were a gift from C. Iacobuzio-Donahue. KPC 4662 cells engineered to express OVA were previously described²⁴ (a gift from R. H. Vonderheide). All cell lines were authenticated as bona fide PDAC cell lines based on histopathologic verification by a dedicated pancreatic cancer pathologist. Orthotopic tumours established with KPC 4662 cells were *Il33*^{High} and transiently decreased in size in response to anti-PD-1 therapy initiated at the time of implantation (anti-PD-1 partial sensitivity). Orthotopic tumours established with KPC 52 cells were *Il33*^{Low} and did not decrease in size in response to anti-PD-1 therapy initiated at the time of implantation (anti-PD-1 resistant). All cell lines were regularly tested using the MycoAlert Mycoplasma Detection Kit (Lonza). Orthotopic PDAC tumours were established as previously described⁶. In brief, mice were anaesthetized using a ketamine–xylazine cocktail, and a small (7-mm) incision was made into the left abdominal side. Tumour cells (10⁶ KPC cells per mouse; 1.25 × 10⁵ KPC-OVA cells per mouse) were suspended in Matrigel (Becton Dickinson), diluted 1:1 with cold phosphate-buffered saline (PBS) (total volume of 50 μl), and injected into the tail of the pancreas using a 26-gauge needle. Successful injection was verified by the appearance of a fluid bubble without intraperitoneal leakage. The abdominal wall was closed with absorbable Vicryl RAPIDE sutures (Ethicon), and the skin was closed with wound clips (Roboz). For subcutaneous PDAC tumours, tumour cells (10⁶ KPC cells per mouse; 1.25 × 10⁵ KPC-OVA cells per mouse) were resuspended in sterile PBS (Fisher Scientific) and implanted subcutaneously. Mice were euthanized at the indicated time points and processed for histology or flow cytometry. Autochthonous KPC mice were euthanized when tumours were detectable by ultrasound. Tumour volumes were measured using serial ultrasound (Vevo 2100 Linear Array Imaging and Vivo LABORATORY Version 3.1.1, Fuji Film Visual Sonics) for orthotopic tumours as previously described²⁵. For subcutaneous tumours, tumour length and width were measured every 2–3 days using calipers, and tumour volumes were calculated as volume = length/2 × width². For survival analyses, survival was determined by a tumour volume of ≥500 mm³ or mouse health requiring euthanasia as defined by institutional IACUC guidelines. No mouse tumours exceeded IACUC-defined maximal tumour volumes of ≥2 cm³. No blinding was performed in experimental mouse interventions, as knowledge of the treatment groups was required.

T cell depletion

CD4⁺ and CD8⁺ cells were depleted by intraperitoneal (i.p.) injection of 250 μg of anti-mouse CD4 antibody (clone GK1.5, Bio X Cell, InVivoPlus)

and 250 μg of anti-mouse CD8a antibody (clone 2.43, Bio X Cell, InVivoPlus). Control mice were treated with rat IgG2b isotype control (clone LTF-2, Bio X Cell, InVivoPlus). Mice were treated daily for 3 days before tumour implantation, and then every 3 days for the duration of the experiment. CD4⁺ and CD8⁺ T cell depletion was confirmed by flow cytometric analysis of tumours and secondary lymphoid organs (>85% depletion).

ILC depletion

ILCs were depleted in *Rag2*^{-/-} mice by i.p. injection of 300 μg of anti-mouse CD90.2 (clone 30-H12, Bio X Cell) on days 0, 1, 3, 6, 9, and 13 following tumour implantation as previously described²⁶. ILC2s were depleted in *Cd4*^{Cre/+} *Icos*^{fl-DTR/+} experimental mice and *Cd4*^{Cre/+} *Icos*^{+/-} control mice treated by i.p. injection of diphtheria toxin (Sigma-Aldrich) at a dosage of 25 ng per gram of mouse body weight. Mice were treated the day before tumour implantation and then every other day thereafter for a total of five doses, as previously described¹⁷. ILC2 depletion was confirmed by flow cytometric analysis of tumours (Extended Data Fig. 5a).

Bone marrow chimaeras

Bone marrow was removed from CD45.2 congenically labelled donor mice, filtered through a 70-mm filter, centrifuged, and resuspended in sterile PBS to a concentration of 10⁸ live cells per 200 μl. CD45.1 congenically labelled C57BL/6J recipient mice were irradiated (5.5 Gy × 2, 6 h apart) 24 h before bone marrow transplant and were maintained on endofloxacin water for 4 weeks after irradiation. A single-cell suspension of CD45.2 bone marrow chimaera in sterile PBS (10⁸ live cells per recipient mouse) was transplanted to each recipient mouse by retroorbital injection. Reconstitution was confirmed by flow cytometry of the peripheral blood at 4 and 8 weeks post transplantation. Tumour implantation experiments were performed at 12 weeks post transplantation.

Recombinant IL33, IL18, and PD-1 blockade

For rIL33, mice were treated with i.p. injections of 500 ng of carrier-free recombinant mouse IL33 (R&D Systems) in sterile PBS daily for 7 days, and then every 2 days thereafter as previously described¹⁶. For rIL18, mice were treated with i.p. injection of 2 μg of carrier-free recombinant mouse IL-18 (R&D Systems) in sterile PBS at days 3, 7, 11, and 15 after tumour inoculation as previously described²⁷. The chimaeric anti-mouse PD-1 antibody (4H2) used in this study was engineered as a mouse IgG1 isotype monoclonal antibody (mAb) and was shown to bind to CHO transfectants expressing PD-1 and to block binding of PD-L1 and PD-L2 to these cells. The affinity of 4H2 for mouse PD-1, determined by surface plasmon resonance using PD-1-Fc, was 4.68 × 10⁻⁹ M. The antibody was produced and purified at Bristol Myers Squibb (BMS). Each batch was certified to have <0.5 EU/mg endotoxin and be of >95% purity. All dosing solutions were prepared in PBS. Mice were treated with i.p. injection of 250 μg anti-PD-1 every 2 days. Transient reduction in tumour size but subsequent regrowth while on continuous anti-PD-1 treatment was defined as a partial response. No reduction in tumour size while on continuous anti-PD-1 was defined as resistance.

Human samples

All tissues were collected at MSKCC following study protocol approval by the MSKCC Institutional Review Board. Informed consent was obtained from all patients. The study was performed in strict compliance with all institutional ethical regulations. All tumour samples were surgically resected primary PDACs.

Tissue microarray. Tissue microarrays (TMAs) were constructed from tumour and adjacent non-tumour cores from formalin-fixed, paraffin-embedded tissue blocks from short-term survivors (*n* = 45 tumours, 5 normal tissues) and long-term survivors (*n* = 51 tumours, 5 normal tissues) of PDAC as previously described⁴. Patient subsets were randomly

selected to undergo tissue microarray construction. Patients treated with neoadjuvant therapy were excluded. All tumours were subjected to pathological re-review and histological confirmation by two expert PDAC pathologists before analysis. Long-term survivors were defined as patients with overall survival of >3 years from surgery and short-term survivors as patients with survival >3 months and <1 year from surgery, to exclude perioperative mortalities. ILC2^{High} and ILC2^{Low} were defined as greater or lesser, respectively, than the median ILC2 frequency for the entire TMA cohort.

Tumour transcriptomic profiling. Patient subsets were randomly selected to undergo transcriptomic profiling as previously described⁴. Patients in the TMA cohort with tumour tissue available for transcriptomic assessment were included in analyses in Fig. 1b to allow protein confirmation of RNA expression. Extracted RNA was qualified on an Agilent BioAnalyzer and quantified by fluorometry (Ribogreen). Preparation of RNA for whole-transcriptome expression analysis was done using the WT Pico Reagent Kit (Affymetrix). Reverse transcription was initiated at the poly-A tail as well as throughout the entire length of RNA to capture both coding and multiple forms of non-coding RNA. RNA amplification was achieved using low-cycle PCR followed by linear amplification using T7 in vitro transcription technology. The cRNA was then converted to biotinylated sense-strand DNA hybridization targets. The prepared target was hybridized to GeneChip Human Transcriptome Array 2.0 (Affymetrix). Washes were performed using the GeneChip Hybridization, Wash and Stain Kit using a Fluidics Station 450/250. Arrays were scanned using the GeneChip Scanner 3000. Data analysis for the array was done using Affymetrix Expression Console Software (SST-RMA algorithm to summarize the signal from array probesets). Immune cytolytic activity was determined as previously described²⁸.

Cell isolation

Mouse and human PDAC tumours and adjacent pancreata were mechanically dissociated and incubated in collagenase (collagenase II for mouse tumours, collagenase IV for human tumours, both 5 mg/ml; Worthington Biochemical Corp., Fisher Scientific), DNase I (0.5 mg/ml; Roche Diagnostics), and Hank's balanced salt solution (Gibco, Fisher Scientific) for 30 min at 37 °C. Digestion was then quenched with fetal bovine serum (FBS, Life Technologies), and cells were filtered sequentially through 100- and 40-µm nylon cell strainers (Falcon, Fisher Scientific). Lymph nodes were mechanically disassociated and filtered through 100- and 40-µm nylon cell strainers (Falcon, Fisher Scientific) using PBS with 1% FBS (Life Technologies). Spleens were mechanically dissociated and filtered through 70- and 40-µm nylon cell strainers (Falcon, Fisher Scientific) using PBS with 1% FBS, followed by RBC lysis (RBC lysis buffer, Thermo Fisher Scientific). Mouse Fc receptors were blocked with FcεRIII/II-specific antibody (1 µg per 1 × 10⁶ cells; clone 2.4G2, Bio X Cell).

ILC2 adoptive transfer

CD45.1 C57BL/6 or *Pdcd1*^{-/-} orthotopic PDAC mice were treated with 500 ng of carrier-free recombinant mouse IL33 (R&D Systems) in sterile PBS daily for 10 days. Live, CD45⁺, lineage⁻, CD90⁺, CD25⁺, ST2⁺ TILC2s were sort-purified to 98% purity at day 10 post-implantation using an Aria Cell sorter (BD Biosciences). TILC2s (5 × 10⁵ cells) were immediately transferred to orthotopic PDAC tumour-bearing *Rora*^{fl/fl} *IL7r*^{Cre/+} CD45.2 mice on days 7 and 14 post-tumour implantation via i.p. injection. Control mice received equivalent volumes of PBS via i.p. injection. Anti-PD-1 treatment in recipient mice was initiated on the day of ILC2 cell transfer. Tissues were collected at indicated time points.

Flow cytometry

Single-cell suspensions were stained using antibody cocktails in the dark at 4 °C, washed, and analysed on a FACS LSR Fortessa (BD

Biosciences). Mouse ILCs were defined as live, CD45⁺, lineage⁻ (CD3, CD5, NK1.1, CD11b, CD11c, CD19, FcεR1), CD25⁺, CD127⁺ cells, as previously described^{2,8}. Mouse immune cells were defined as follows: ILC2s: live, CD45⁺, lineage⁻, CD25⁺, ST2⁺ cells; central memory T cells (T_{CM}): live, CD45⁺, CD3⁺, NK1.1⁻, CD8⁺, CD62L⁺, CD44⁺; dendritic cells (DC): live, CD45⁺, CD3⁺, NK1.1⁻, Gr1⁻, F4/80⁺, CD11c⁺, MHC-II⁺; B cells: live, CD45⁺, CD3⁺, CD19⁺; T cells: live, CD45⁺, CD3⁺; CD4⁺ T cells: live, CD45⁺, CD3⁺, CD4⁺; CD8⁺ T cells: live, CD45⁺, CD3⁺, CD8⁺; regulatory T cells: live, CD45⁺, CD3⁺, CD4⁺, FoxP3⁺; tumour-associated macrophages: live, CD45⁺, CD11b⁺, F4/80⁺, Gr1⁺; myeloid-derived suppressor cells (MDSCs): live, CD45⁺, CD3⁺, CD11b⁺, F4/80⁺, Gr1⁺. Mouse cells were stained with the following antibodies: from Biolegend, CD45 (clone 30-F11, Pacific Blue), CD45.1 (clone A20, BV711), NK1.1 (clone PK136, APC), Gr-1 (clone RB6-8C5, BV605), CD103 (clone 2E7, BV711); from BD Biosciences, CD5 (clone 53-7.3, APC), CD11c (clone HL3, APC), NK1.1 (clone PK136, BV605), CD4 (clone RM4-5, BV786), CD62L (clone MEL-14, APC), CD19 (clone 1D3, BV510), Ly6C (clone AL-21, PerCP-Cy5.5), Ly6G (clone 1A8, AF700), PD-1 (clone J43 BV605), TNF-α (clone MP6-XT22, BV510), IFN-γ (clone XMG1.2, APC-Cy7), CD90.2 (clone 53-2.1, BV786), T-bet (clone Q4-46, BV711), RORγ-t (clone Q31-378, BV786), GATA3 (clone L50-823, PE-Cy7), and IL4 (clone 11B11, BV650); from Thermo Fisher Scientific CD3 (clone 17A2, Alexa Fluor 700), CD11b (clone M1/70, APC), CD11b (clone M1/70, PerCP-Cy5.5), CD8 (clone 53-6.7, Alexa Fluor 700), CD19 (clone 1D3, Alexa Fluor 700), FcεR1 (clone MAR-1, APC), F4/80 (clone BM8, PE-Cy5), CD3 (clone 145-2C11, PE-Cy7), MHC-II (clone M5/114.15.2, Alexa Fluor 700), CD44 (clone IM7, PerCP-Cy5.5), CD127 (clone A7R34, FITC), CD25 (clone PC61.5, PerCP-Cy5.5), IL5 (clone TRFK5, PE), CD11c (clone N418, FITC), ST2 (clone RMST2-2, PE-Cy7), and FOXP3 (clone FJK-16S, APC); and from MBL international, SINFEKL tetramer (catalogue # TB-5001-1, PE).

Human ILCs were defined as live, CD45⁺, lineage⁻ (CD3, CD5, CD56, CD11b, CD11c, CD16, CD19, TCRα/β, FcεR1), CD25⁺, CD127⁺ cells as previously described⁸. Human cells were stained with the following antibodies: from BD Biosciences, GATA3 (clone L50-823, BV711), T-bet (clone O4-46, BV650), RORγ-T (clone Q21-559, PE); from Biolegend, CRTH2 (clone BMI6, PE-Cy7), CD11b (clone ICRF44, APC), CD56 (clone NCAM16.2, BV650), CD25 (clone BC96, PerCP-Cy5.5), CD45 (clone HI30, Pacific Blue), TCRα/β (clone IP26, APC); from Thermo Fisher Scientific, CD16 (clone CB16, APC), CD11c (clone 3.9, APC), CD127 (clone RDR5, FITC), CD3 (clone OKT3, Alexa Fluor 700), ST2 (clone hIL33Rcap, PE), CD5 (clone L17F12, APC), CD19 (clone HIB19, AF700), FcεR1 (clone AER-37, APC). All samples for flow cytometry were prospectively collected from unselected patients with PDAC.

To examine intracellular cytokine production, single-cell suspensions of tumours were stimulated for 6 h ex vivo with phorbol 12-myristate (PMA, 100 ng/ml) and ionomycin (1 ng/ml) in the presence of brefeldin A (10 µg/ml) (all from Sigma-Aldrich) at 37 °C. Cells were then surface-stained, fixed, permeabilized, and stained for cytokine production using the Fixation and Permeabilization Buffer Kit per the manufacturer's recommendations (Invitrogen, Thermo Fisher Scientific). Appropriate isotype controls were used as indicated. Analysis was performed using FlowJo (versions 9 and 10, Tree Star).

Immunohistochemistry

Tissues were fixed in paraformaldehyde (Fisher Scientific) for 24 h and embedded in paraffin. The tissue sections were deparaffinized with EZPrep buffer (Ventana Medical Systems), then antigen retrieval was performed with CC1 buffer (Ventana Medical Systems). Sections were blocked for 30 min with Background Buster solution (Innovex), followed by avidin-biotin blocking for 8 min (Ventana Medical Systems). Mouse IL33 (AF3626, R&D Systems), mouse smooth muscle actin (Abcam), and human IL33 (AF3625, R&D Systems) antibodies were applied, and sections were incubated for 4 h, followed by a 60-min incubation with biotinylated rabbit anti-goat IgG (Vector labs), or biotinylated goat

Article

anti-rabbit IgG (Vector labs) at 1:200 dilution. Detection was performed with a DAB detection kit (Ventana Medical Systems) according to the manufacturer's instructions. Any section containing cells demonstrating cytoplasmic or nuclear positivity for IL33 was designated to have positive staining. Slides were counterstained with Masson's trichrome, or haematoxylin and eosin, and coverslipped with Permount (Fisher Scientific). All histologic sections were evaluated by an independent PDAC pathologist.

Mouse immunofluorescence

IL33/CD11b/CK19/Iba1 immunofluorescence. Multiplex immunofluorescent staining was performed using a Discovery XT processor (Ventana Medical Systems) as described²⁹.

IL33. First, sections were incubated with anti-mIL33 (R&D Systems, catalogue # AF3626, 1 µg/ml) for 4 h, followed by 60 min incubation with biotinylated horse anti-goat IgG (Vector Laboratories) at 1:200 dilution. Detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa Fluor 488 (Invitrogen) prepared according to the manufacturer's instructions with predetermined dilutions.

CD11b. Next, sections were incubated with anti-CD11b (Abcam, clone EPR1544) for 5 h, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector Laboratories) at 1:200 dilution. Detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa 594 (Invitrogen) prepared according to the manufacturer's instructions with predetermined dilutions.

CK19. Next, slides were incubated with anti-CK19 (Abcam, clone EP1580Y) for 5 h, followed by 60 min incubation with biotinylated goat anti-rabbit (Vector Laboratories) at 1:200 dilution. Detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa Fluor 546 (Invitrogen) prepared according to the manufacturer's instructions with predetermined dilutions.

Iba1. Finally, sections were incubated with anti-Iba1 (Wako, catalogue #019-19741) for 5 h, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector Laboratories) at 1:200 dilution. Detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa 647 (Invitrogen) prepared according to the manufacturer's instructions with predetermined dilutions. After staining, slides were counterstained with DAPI (Sigma-Aldrich) for 10 min and coverslipped with Mowiol.

Human immunofluorescence

Tissue sections were deparaffinized with proprietary Leica Bond buffer (Leica Biosystems), and antigen retrieval was performed with Leica Bond ER2 buffer (Leica Biosystems). First, sections were incubated with anti-PD-1 (Cell Marque, clone NAT105) for 1 h, followed by detection with Bond Polymer Refine Detection kit (Leica Biosystems) and Tyramide Alexa Fluor 488 (Invitrogen). Next, sections were incubated with anti-CD3 (DAKO, catalogue #A0452) for 1 h, followed by detection with Bond Polymer Refine Detection kit (Leica Biosystems) and Tyramide CF594 (Biotum). Next, sections were incubated with anti-GATA3 (Cell Marque, clone L50-823) for 1 h, followed by detection with Bond Polymer Refine Detection kit (Leica Biosystems) and CF 543 (Biotum). Finally, sections were incubated with anti-CD45 (DAKO, clone 2B11 + PD7/26) for 1 h, followed by detection with Bond Polymer Refine Detection kit (Leica Biosystems) and Tyramide Alexa Fluor 647 (Invitrogen). All detections were prepared according to the manufacturer's instructions with predetermined dilutions. After staining, slides were counterstained with DAPI (Sigma-Aldrich) for 10 min and coverslipped with Mowiol.

Digital image processing and analysis

Slides were digitized using Panoramic Flash 250 (3Dhistech, Budapest, Hungary) using a Zeiss 20×/0.8NA objective and custom filters for A488, A546, A594, and A647. Each core was exported into a multi-channel tiff file and analysed using a custom macro written in Fiji/ImageJ. For quantification, each nucleus was segmented using the DAPI channel after appropriate processing and background subtraction. Then for each nucleated cell, the presence or absence of the other markers was assessed after setting appropriate thresholds for each marker. The number of cells with specific combinations of markers was tallied. ILC2s were defined as CD45⁺ CD3⁻ GATA3⁺ nucleated cells, PD-1-expressing ILC2s were defined as CD45⁺ CD3⁻ GATA3⁺ PD-1⁺ nucleated cells, and PD-1-expressing T cells were defined as CD45⁺ CD3⁺ PD-1⁺ nucleated cells. For each patient, the frequency of each cell type as a fraction of all nucleated cells was calculated in triplicate cores, followed by determination of the mean frequency of triplicate cores to calculate the final cellular frequency per patient.

RNA sequencing

Mouse. Tissues from mice with orthotopic PDACs ($n = 6$) were collected and dissociated into single-cell suspensions as described above. Tumour-infiltrating leukocytes were positively selected by magnetically activated cell sorting using mouse CD45 MicroBeads (Miltenyi Biotec). Purification of magnetically activated sorted cells was confirmed by flow cytometry and was >95%. RNA was isolated from the sorted cells using an RNeasy Plus Mini Kit (Qiagen). Poly(A) capture and paired-end RNA-seq were performed by the MSKCC Integrated Genomics Core Facility. Specifically, after RiboGreen quantification and quality control by Agilent BioAnalyzer, 500 ng of total RNA underwent polyA selection and TruSeq library preparation according to instructions provided by Illumina (TruSeq Stranded mRNA LT Kit, catalogue # RS-122-2102), with eight cycles of PCR. Samples were barcoded and run on a HiSeq 4000 in a 100 bp/100 bp paired-end run, using the HiSeq 3000/4000 SBS Kit (Illumina). An average of 83 million paired reads was generated per sample. Ribosomal reads represented at most 0.03% of the total reads generated, and the percentage of mRNA bases averaged 76.6%. The expression data set was loaded into Gene Set Enrichment Analysis (GSEA) 3.0. Gene set databases for antigen presentation and T cell mediated immunity were selected from MSIGDB v6.1, with a false discovery rate of ≤ 0.25 to facilitate exploratory discovery. GSEA was run with 1,000 permutations. Three gene set databases met this threshold: GO: 0002474 (antigen processing and presentation of peptide antigen via MHC class I); GO: 0002711 (positive regulation of t cell mediated immunity); and GSE19825 (naive vs day 3 effector CD8 T cell up).

Single-cell RNA sequencing

Library preparation for single-cell immune profiling, sequencing, and post-processing of the raw data were performed at the Epigenomics Core at Weill Cornell Medicine.

Single-cell RNA library preparation and sequencing. Single-cell suspensions of fluorescence-activated cell (FAC)-sorted ILC2 cells from pancreatic KPC tumours and mesenteric DLNs from mice treated with vehicle, rIL33 alone, or rIL33 + anti-PD-1 were prepared as described above. scRNA-seq libraries were prepared according to 10X Genomics specifications (Chromium Single Cell V(D)J User Guide PN-1000006, 10X Genomics). Four independent cellular suspensions (85–90% viable) at a concentration between 90 and 200 cells/µl were loaded onto the 10X Genomics Chromium platform to generate Gel Beads-in-Emulsion (GEM), targeting about 2,000 single cells per sample. After GEM generation, the samples were incubated at 53 °C for 45 min in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad) to generate polyA cDNA barcoded at the 5' end by the addition of a template switch oligo (TSO) linked to a cell barcode and Unique

Molecular Identifiers (UMIs). GEMs were broken, and the single-strand cDNA was cleaned up with DynaBeads MyOne Silane Beads (Thermo Fisher Scientific). The cDNA was amplified for 16 cycles (98 °C for 45 s; 98 °C for 20 s, 67 °C for 30 s, 72 °C for 1 h). The quality of the cDNA was assessed using an Agilent Bioanalyzer 2100, obtaining a product of about 1,200 bp. Fifty nanograms of cDNA was enzymatically fragmented, end repaired, A-tailed, subjected to a double-sided size selection with SPRIselect beads (Beckman Coulter), and ligated to adaptors provided in the kit. A unique sample index for each library was introduced through 14 cycles of PCR amplification using the indexes provided in the kit (98 °C for 45 s; 98 °C for 20 s, 54 °C for 30 s, 72 °C for 20 s × 14 cycles; 72 °C for 1 min; held at 4 °C). Indexed libraries were subjected to a second double-sided size selection, and libraries were then quantified using Qubit fluorometric quantification (Thermo Fisher Scientific). The quality was assessed on an Agilent Bioanalyzer 2100, obtaining an average library size of 450 bp. No treatment samples had concentrations below detectable limits, and cDNA amplification was done with 18 cycles and sample index with 16 cycles. Libraries were diluted to 10 nM and clustered using a NovaSeq600 on a paired-end read flow cell and sequenced for 28 cycles on R1 (10X barcode and the UMIs), followed by 8 cycles of 17 Index (sample Index), and 89 bases on R2 (transcript), obtaining about 100 million clusters per sample, except for tumours from vehicle-treated mice (clustered at about 10 million). Primary processing of sequencing images was done using Illumina's Real Time Analysis software (RTA). 10X Genomics Cell Ranger Single Cell Software suite v3.0.2 (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger>) was used to perform sample demultiplexing, alignment to mouse genomic reference mm10, filtering, UMI counting, single-cell 5' end gene counting, and quality control using the manufacturer's parameters. Data from approximately 11,000 single cells that passed quality control were obtained with approximately 41,000 mean reads per cell (48% sequencing saturation).

scRNA-seq data processing. The Seurat R package version 3.1 pipeline was used to identify clusters on combined data sets³⁰. First, individual data sets were read into R as count matrices and converted into Seurat objects, selecting on genes expressed in ≥3 cells and on cells with at least 200 detected genes. A standard pre-processing workflow was then used to filter cells based on excluding cells with either more than 2,500 or fewer than 200 unique genes expressed, and cells with greater than 5% mitochondrial gene content.

Following filtering, the samples were merged, and the gene expression measurements for retained cells were log-transformed, normalized by total expression per cell, and scaled to 10,000 molecules per cell. The top 2,000 highly variable genes across the single cells were then identified, and principal component (PC) analysis was conducted. After examining jackstraw and elbow plots, we selected the top 15 PCs for clustering using *K*-nearest neighbour (KNN) clustering with cluster resolution set at 0.4, identifying 6–8 clusters in all samples-combined and tumour-combined merged data sets. Nonlinear dimensional reduction with UMAP was used to visualize the data sets, also using the top 15 PCs. Differential gene expression for gene marker discovery across the clusters was performed using the Wilcoxon rank sum test as used in the Seurat package. Pairwise comparison using Wilcoxon rank sum test was performed with the Holm *P* value adjustment method to compare gene expression between samples.

In vitro assays

KPC 4662 cells were cultured for 1 week in a 96-well flat-bottomed plate (Falcon) in complete medium: RPMI with L-glutamine (Gibco, Thermo Fisher Scientific) with 10% fetal bovine serum (Life Technologies), 100 units/ml of penicillin, 100 µg/ml of streptomycin, and rIL33 at concentrations of 0, 10, 100, and 500 ng/ml. Culture medium and cytokines were replenished every 48 h. Viability was measured using

a colorimetric tetrazolium salt assay (Cell Counting Kit, Dojindo Molecular Technologies) per the manufacturer's instructions and read on a Synergy HT Multi-Detection Microplate Reader (Biotek). Cells were collected and stained for Annexin V (Thermo Fisher Scientific), Ki-67 (clone SolA15, Thermo Fisher Scientific), and ST2 (clone RMST2, Thermo Fisher Scientific). For all in vitro experiments, 2–3 technical replicates were performed per independent experiment.

In vitro dendritic cell migration assays. Mouse splenic DCs were isolated and enriched using a mouse pan DC isolation kit according to the manufacturer's protocol (Miltenyi Biotec). Flow cytometry was used to assess DC purity (>70% CD11c⁺ of live cells). Cells were plated in complete RPMI medium at 5 × 10⁵ cells/ml with 50 ng/ml of recombinant mouse GM-CSF (Biolegend) overnight. Next, chemotaxis of splenic DCs was analysed by transwell migration assays. RPMI (600 µl) with or without 100 ng/ml of recombinant mouse CCL5 (Biolegend) was added to the lower chambers of a 6.5-mm Transwell plate with 5.0-µm pore polycarbonate membrane inserts (Sigma-Aldrich). RPMI (200 µl) was also added to the upper chambers, and plates were allowed to equilibrate at 37 °C in 5% CO₂ for 15 min. Splenic DCs (1 × 10⁵ cells in 100 µl RPMI) were then loaded into the upper chambers and incubated at 37 °C in 5% CO₂ for 2 h. After incubation, membrane inserts were carefully removed, and cells were collected from the lower chambers. Migrated DCs were incubated with DAPI and CD11c antibodies for 20 min at 4 °C, and Precision Count Beads (Biolegend) were added to quantify the number of live migrated CD11c⁺ cells using flow cytometry according to the manufacturer's protocol.

Statistics

Data are expressed as median. As we observed many statistically significant effects in the data without a priori sample size calculations, no statistical methods were used to determine sample size. Comparisons between two groups were performed using unpaired Mann–Whitney test with the Benjamini–Krieger–Yekutieli false discovery approach for multiple time point comparisons (two-tailed). Comparisons among multiple groups were performed using one-way ANOVA test followed by Kruskal–Wallis multiple comparison post-test. Comparisons among multiple groups across multiple time points were performed using two-way ANOVA test. Correlations between two variables were calculated using linear regression. Survival curves were compared by two-sided log-rank test. Tumour incidences were compared by χ^2 test. All alpha levels were 0.05, with *P* < 0.05 considered a significant difference. Statistical analyses were performed using Prism 7.0 (GraphPad Software).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source code for immune quantification is available in Supplementary Data 1. Bulk RNA-seq data are available under Gene Expression Omnibus (GEO) accession number GSE129388. scRNA-seq data are available under GEO accession number GSE136720. Source data are provided for all experiments. All other data are available from the corresponding author upon reasonable request.

23. Donovan, C. et al. Roles for T/B lymphocytes and ILC2s in experimental chronic obstructive pulmonary disease. *J. Leukoc. Biol.* **105**, 143–150 (2019).
24. Evans, R. A. et al. Lack of immunoeediting in murine pancreatic cancer reversed with neoantigen. *JCI Insight* **1**, 88328 (2016).
25. Sastra, S. A. & Olive, K. P. Quantification of murine pancreatic tumors by high-resolution ultrasound. *Methods Mol. Biol.* **980**, 249–266 (2013).
26. Monticelli, L. A. et al. Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nat. Immunol.* **12**, 1045–1054 (2011).
27. Ma, Z. et al. Augmentation of immune checkpoint cancer immunotherapy with IL18. *Clin. Cancer Res.* **22**, 2969–2980 (2016).

28. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

29. Yarilin, D. et al. Machine-based method for multiplex in situ molecular characterization of tissues by immunofluorescence detection. *Sci. Rep.* **5**, 9534 (2015).

30. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

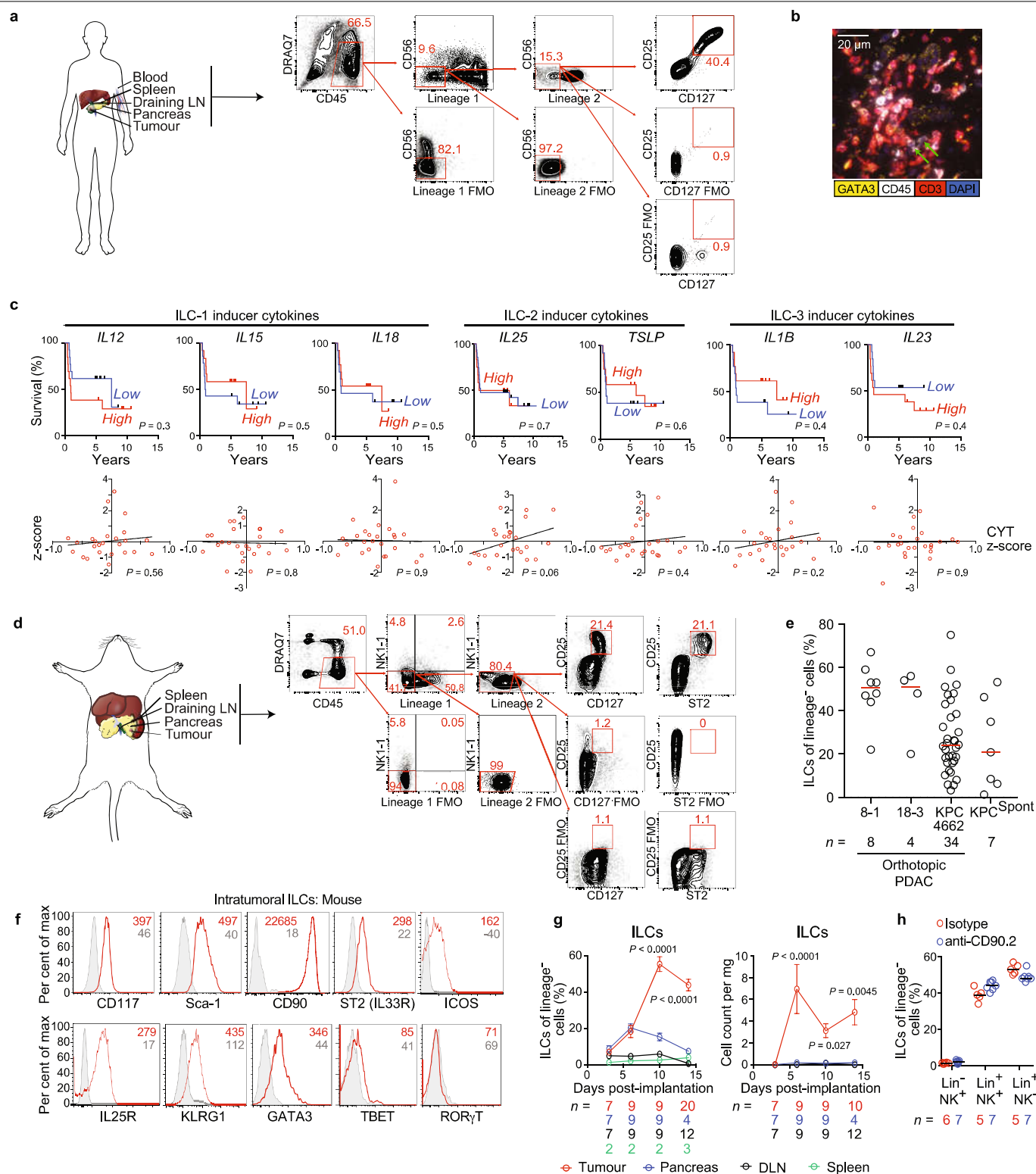
Acknowledgements We thank J. Novak, J. Moore and E. Patterson for editorial assistance; B. Medina, G. Vitiello, J. Zhang, S. Zeng, F. Rossi, J. Loo, N. Param, J. Maltbaek, O. Grbovic-Huezo, Y. Senbabaoglu, M. Gigoux, R. Giese and S. Budhu for helpful discussions and technical assistance; and the Epigenomics Core of Weill Cornell Medical College for technical assistance with scRNA-seq. This work was supported by the V Foundation Convergence Scholar Grant (J.A.M., J.D.W., V.P.B.), the Stand Up to Cancer Convergence Award (J.D.W., V.P.B.), the National Cancer Institute K12CA184746-01A1 (V.P.B.), the Damon Runyon Clinical Investigator Award (V.P.B.), the Ben and Rose Cole Pria Foundation Scholar Award (V.P.B.), the Sarah Min and Matthew Pincus Pancreatic Cancer Immunotherapy Award (V.P.B.), an administrative supplement to NIH P30-CA008748 (S.D.L., V.P.B.), NIH R01 CA204228, NIH P30CA023108 (S.D.L.), Swim Across America and the Ludwig Institute for Cancer Research (J.D.W., T.M.), and the Parker Institute for Cancer Immunotherapy (J.D.W., T.M.). Services by the MSKCC Small-Animal Core Facility and Integrated Genomics Core were funded by the National Cancer Institute Cancer Center Support Grant (P30 CA008748-48), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

Author contributions V.P.B. conceived the study. J.A.M, J.L., L.A.R., S.D.L., J.D.W., R.P.D., T.M. and V.P.B. designed all the experiments. J.A.M, J.L., L.A.R., J.R., J.Z., A.R. and V.P.B. performed all the experiments. B.G. assisted with generation of bone marrow chimaeras. J.L., Z.S. and D.R. analysed the scRNA-seq results. G.A. and U.B. performed the pathologic analyses. E.E., Y.P. and D.A.T. generated and assisted in experiments on autochthonous KPC mice. M. Gururajan

provided technical assistance with the PD-1 blocking antibody. M. Gönen provided statistical oversight. J.A.M., J.L., L.A.R., J.R., S.D.L., R.P.D., T.M. and V.P.B. analysed all the data. All authors interpreted the data. J.A.M. and V.P.B. wrote the manuscript with input from all authors.

Competing interests V.P.B. is a recipient of an immuno-oncology translational research grant from Bristol-Myers Squibb and is an inventor on a patent application related to work on neoantigen modelling. S.D.L. is a member of the scientific advisory board of Nybo Pharmaceuticals, and co-founder of Episteme Prognostics. J.D.W. is a consultant for Adaptive Biotech, Advaxis, Amgen, Apricity, Array BioPharma, Ascentage Pharma, Astellas, Bayer, Beigene, Bristol-Myers Squibb, Celgene, Chugai, Elucida, Eli Lilly, F Star, Genentech, Imvaq, Janssen, Kleo Pharma, Linneaus, MedImmune, Merck, Neon Therapeutics, Ono, Polaris Pharma, Polynoma, Psioxus, Puretech, Recepta, Trieza, Sellas Life Sciences, Seramatrix, Surface Oncology and Syndax; is a recipient of research support from Bristol-Myers Squibb, Medimmune, Merck Pharmaceuticals and Genentech; has equity in Potenza Therapeutics, Tizona Pharmaceuticals, Adaptive Biotechnologies, Elucida, Imvaq, Beigene, Trieza and Linneaus; and has received an honorarium from Esanex. T.M. is a consultant for Immunosis Therapeutics and Pfizer; is a co-founder with equity in Imvaq Therapeutics; receives research funding from Bristol-Myers Squibb, Surface Oncology, Kyn Therapeutics, Infinity Pharmaceuticals Inc., Peregrine Pharmaceuticals Inc., Adaptive Biotechnologies, Leap Therapeutics Inc. and Aprea; and is an inventor on patent applications related to work on Oncolytic Viral Therapy, Alpha Virus Based Vaccine, Neo Antigen Modeling, CD40, GITR, OX40, PD-1 and CTLA-4. M. Gururajan is an employee of Bristol-Myers Squibb and has financial interest in the company.

Additional information
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2015-4>.
Correspondence and requests for materials should be addressed to T.M. or V.P.B.
Peer review information *Nature* thanks Richard Locksley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at <http://www.nature.com/reprints>.

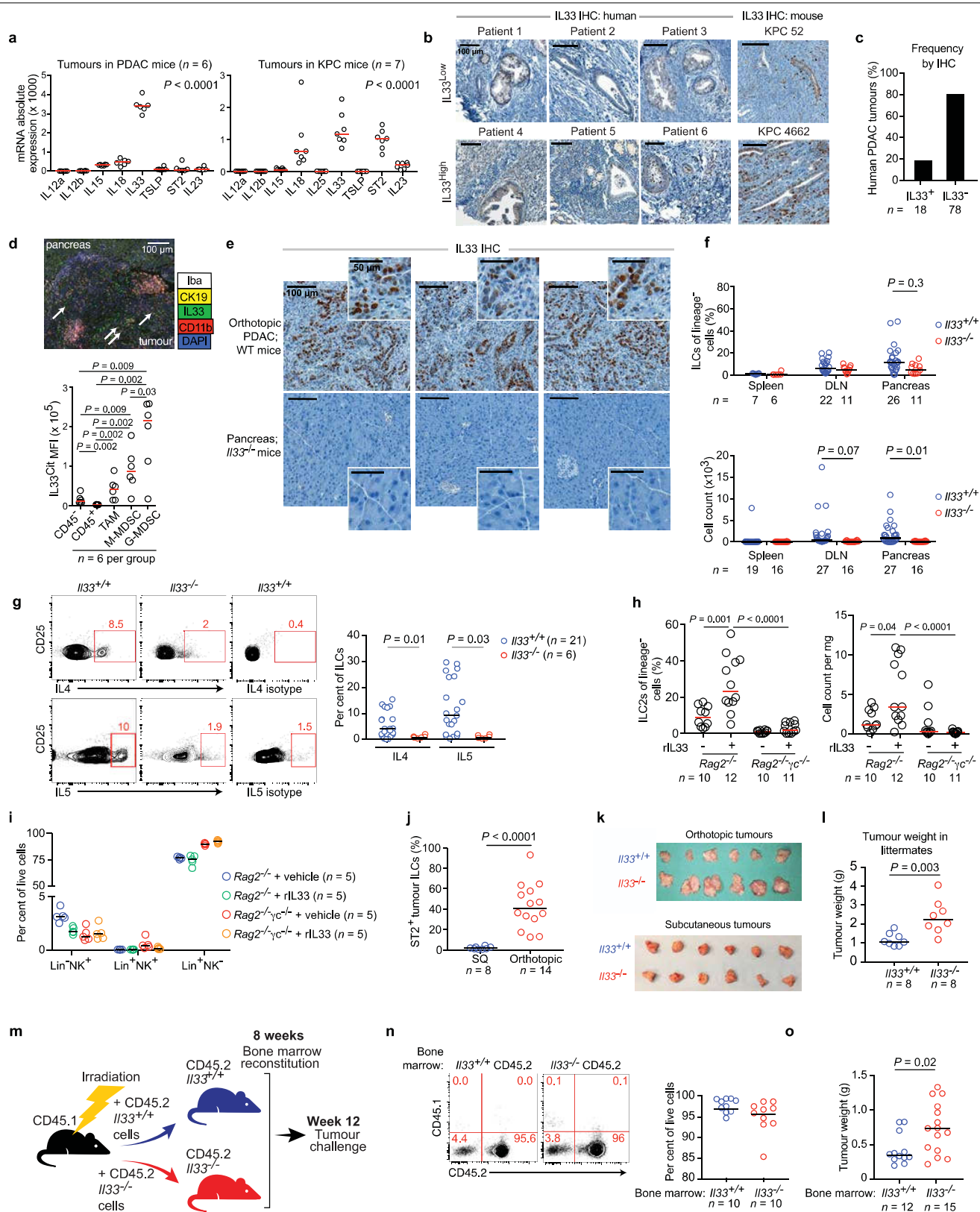


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Identification of IL33-dependent ILCs in pancreatic cancer. **a**, Gating strategy to identify human ILCs. The first plot was pre-gated on live (DRAQ7⁻) cells and singlets. Lineage 1 cocktail: CD5, CD11b, CD11c, CD16, FcεR1. Lineage 2 cocktail: CD3, CD19, TCRα/β. ILCs were identified as lineage⁻ CD56⁻ CD25⁺ CD127⁺ cells. FMO, fluorescence minus one. **b**, Representative image of immunofluorescence of ILC2s in tumour tissue microarrays from short- and long-term PDAC survivors ($n = 96$). Arrows, putative ILC2s. **c**, Top, overall survival of patients with more (high) or less (low) than the median intratumoral mRNA level of ILC-stimulating cytokines. Bottom, correlation between expression of ILC-activating cytokines and immune cytolytic index (CYT) in short- and long-term survivors of human PDAC. Curves were fit by linear regression. $n = 25$. **d**, Gating strategy to identify mouse ILCs. The first plot was pre-gated on live (DRAQ7⁻) cells and singlets. Lineage 1 cocktail: CD5, CD11b, CD11c, FcεR1. Lineage 2 cocktail: CD3, CD19. ILCs were identified as lineage⁻ NK1.1⁻ CD25⁺ CD127⁺, and ILC2s were identified as lineage⁻

NK1.1⁻ CD25⁺ ST2⁺ cells. Gating on orthotopic PDAC mice shown.

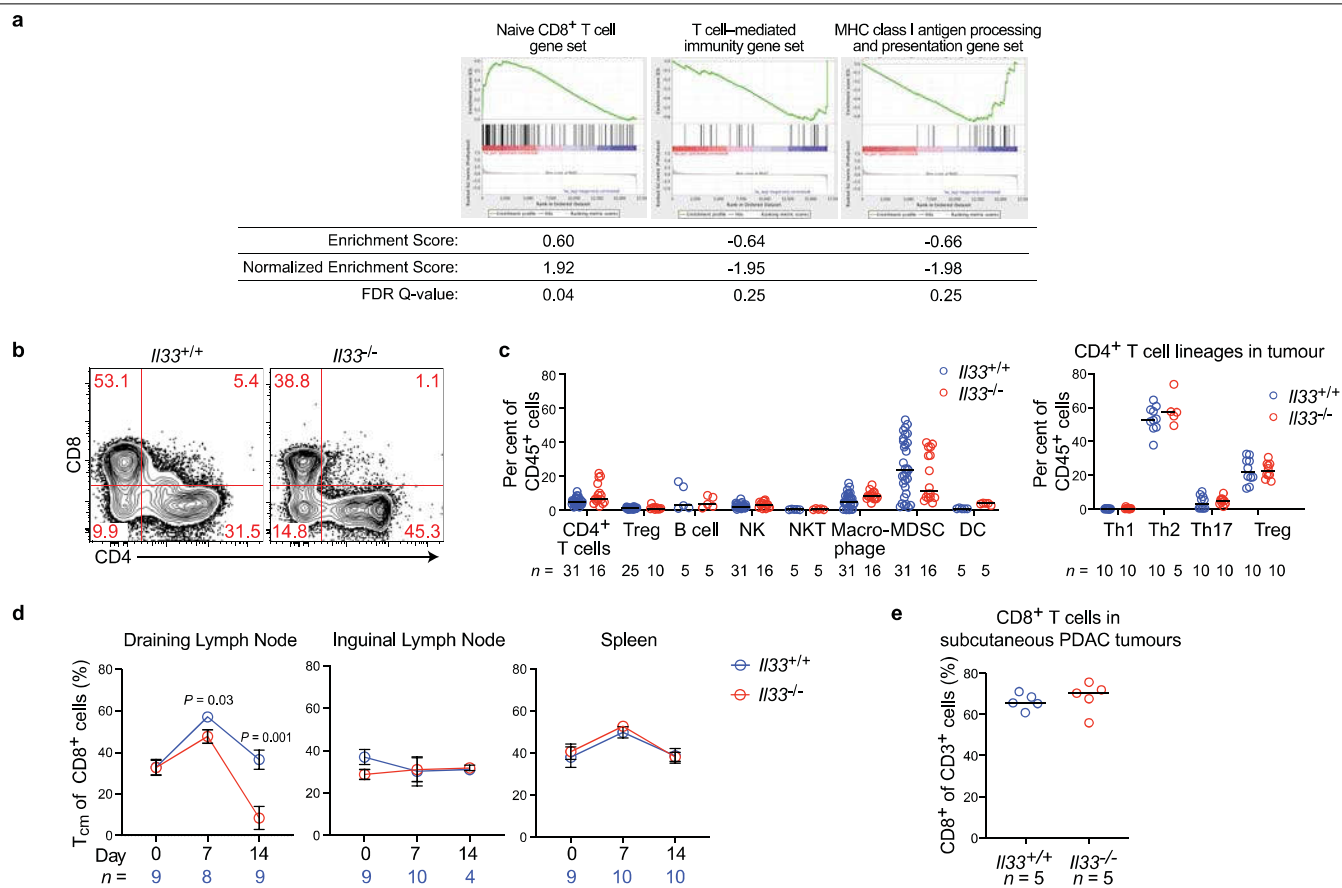
e, Intratumoral ILC frequency in orthotopic PDAC mice established with KPC cell lines 8-1, 18-3, and in autochthonous KPC mice with spontaneous PDAC (KPC^{Spont}). Composite ILC frequencies from Fig. 1d and others are included for comparison (KPC 4662). **f**, Phenotype of ILCs in PDAC mice. Grey curves, isotype controls; numbers, mean fluorescence intensity. **g**, Expansion kinetics of ILCs in tissues from PDAC mice. **h**, Changes in non-ILC cell frequency in *Rag2*^{-/-} PDAC mice treated with anti-CD90.2 or isotype antibodies. Data were collected 14 days (**d–f**), 10 days (**h**), or at the indicated time points after tumour implantation. n indicates individual mice analysed separately in at least two independent experiments with $n \geq 2$ per group. Median \pm s.e.m; horizontal bars show median. P values determined by two-sided log-rank test (**c**, top), linear regression (**c**, bottom), or two-tailed Mann–Whitney test (**g**). P values in **g** indicate tumour comparisons to all other tissues.



Extended Data Fig. 2 | See next page for caption.

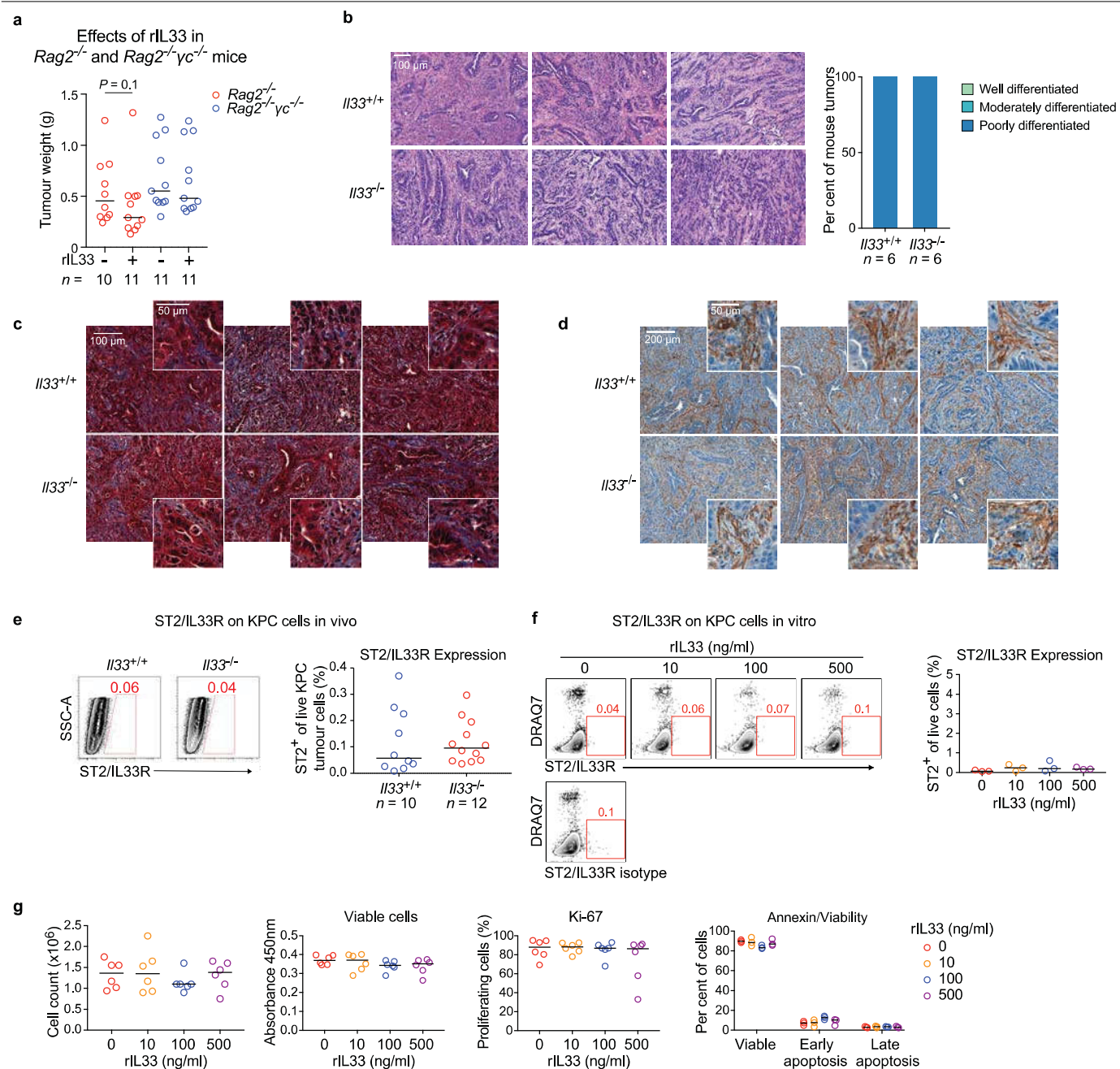
Extended Data Fig. 2 | Host-derived IL33 activates pancreatic ILC2s. **a**, mRNA expression of ILC1- (IL12, IL15, IL18), ILC2- (IL25, IL33, TSLP), and ILC3-inducer cytokines (IL23) and the IL33 receptor (ST2) in orthotopic PDAC tumours (left) and autochthonous PDAC tumours in KPC mice from a previously published mRNA microarray (right)¹². **b**, Representative IL33 immunohistochemistry (IHC) of IL33^{low} and IL33^{high} human (tissue microarray, $n = 96$) and mouse PDAC ($n = 3$ per group). **c**, Frequency of patients with PDAC demonstrating IL33 positivity by IHC in a human PDAC tumour microarray. **d**, Multiplexed immunofluorescence for IL33, ductal marker CK19, and myeloid markers CD11b and Iba in mouse PDAC (top). Arrows, IL33-expressing cells. IL33 mean fluorescence intensity (MFI) in non-immune (CD45⁻), immune (CD45⁺), macrophage (TAM), and monocytic and granulocytic myeloid-derived suppressor cell (M-MDSC and G-MDSC) populations in tumours from IL33^{cit} reporter PDAC mice (bottom). **e**, Representative IL33 protein expression shown by IHC in orthotopic PDAC tumours in IL33^{+/+} (WT) mice, and non-tumour-bearing pancreata in IL33^{-/-} mice ($n = 3$ per group). **f**, ILC frequency (top)

and cell number (bottom) in organs and DLNs of IL33^{+/+} and IL33^{-/-} orthotopic PDAC mice. **g**, Gating and frequency of IL4 and IL5 expression in intratumoral ILCs in IL33^{+/+} and IL33^{-/-} orthotopic PDAC mice. **h, i**, ILC2 (**h**) and immune cell frequencies (**i**) in orthotopic Rag2^{-/-} and Rag2^{-/-}γc^{-/-} PDAC mice with or without treatment with rIL33. **j**, Frequency of ST2⁺ tumour ILCs in mice with subcutaneous (SQ) and orthotopic PDACs. **k**, Tumours in orthotopic and subcutaneous PDAC mice. **l**, Tumour weight in IL33^{+/+} and IL33^{-/-} littermate PDAC mice. **m**, Experimental schema of bone-marrow chimaeras to evaluate contribution of haematopoietic cell-derived IL33 to tumour control. **n, o**, Haematopoietic cell reconstitution (**n**) and tumour weight (**o**) in irradiated CD45.1 congenic mice reconstituted with either CD45.2 IL33^{+/+} or CD45.2 IL33^{-/-} bone marrow. Data were collected 14 (**a, b, d-g, j, k** (orthotopic), **l, o**), 28 (**k** (subcutaneous)), or 10 (**h, i**) days after tumour implantation. Horizontal bars show median. n indicates individual mice analysed separately in at least two independent experiments with $n \geq 2$ per group. P values determined by one-way ANOVA (**a**) or two-tailed Mann-Whitney test (**d, f-h, j, l, o**).



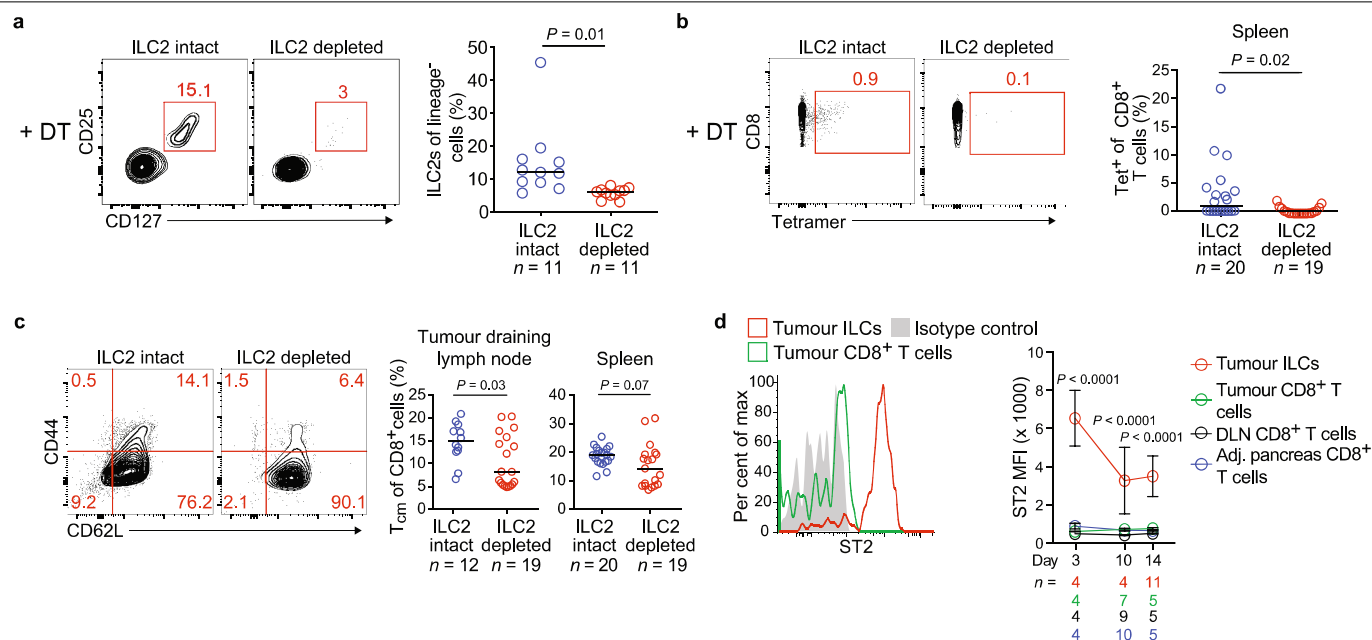
Extended Data Fig. 3 | Host-derived IL33 activates pancreatic T cell immunity. **a**, Gene set enrichment analysis of bulk RNA-seq from purified CD45⁺ immune cells from *Il33*^{+/+} and *Il33*^{-/-} PDAC mice. Enrichment plots and enrichment scores are shown for three gene sets comparing expression in *Il33*^{-/-} to expression in *Il33*^{+/+} mice ($n = 3$ mice per group). FDR, false discovery rate. **b, c**, Gating of CD8⁺ T cells (**b**) and frequencies of various immune cell types (**c**, left) and CD4⁺ T cell lineages (**c**, right) in *Il33*^{+/+} and *Il33*^{-/-} orthotopic PDAC mice. **d**, Frequency of T central memory (T_{CM}) cells (CD45⁺ CD3⁺ CD8⁺ CD44⁺ CD62L⁺) in tumour DLNs and non-tumour draining distant lymphoid

organs (inguinal lymph node and spleen) in *Il33*^{+/+} and *Il33*^{-/-} orthotopic PDAC mice. **e**, Frequency of CD8⁺ T cells in subcutaneous PDAC tumours. NK, natural killer cells; NKT, natural killer T cells; T_{reg} , regulatory T cells; MDSC, myeloid-derived suppressor cells; DC, dendritic cells. Data were collected 14 days after tumour implantation or at the time points indicated. Median \pm s.e.m; horizontal bars show median. n indicates individual mice analysed separately in at least two independent experiments with $n \geq 2$ per group. P values determined by one-way ANOVA (**d**).



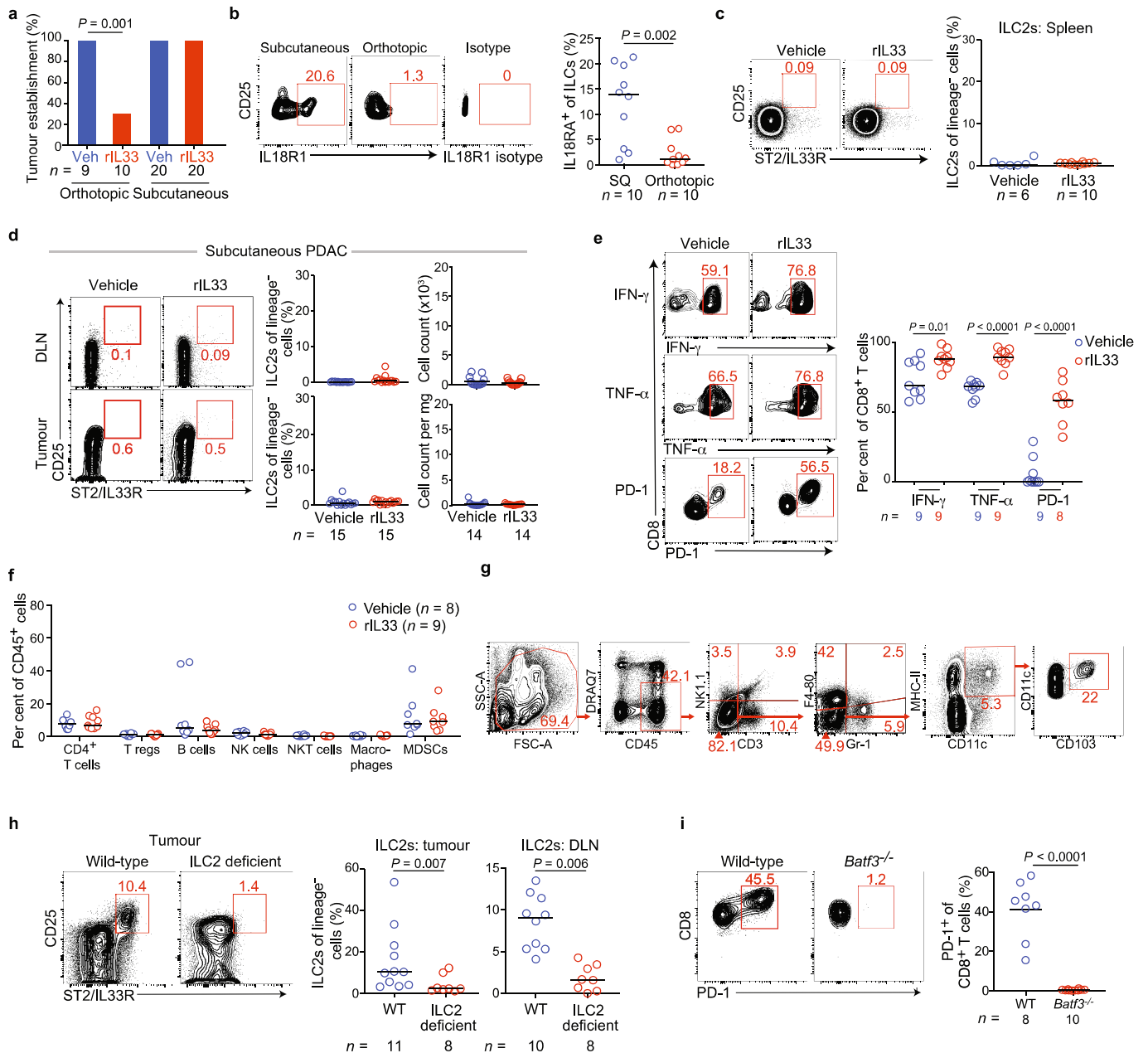
Extended Data Fig. 4 | IL33 and ILCs do not directly induce tumour cell death. a, Tumour weight in *Rag2^{-/-}* and *Rag2^{-/-}γc^{-/-}* PDAC mice treated with vehicle or rIL33. **b**, Representative haematoxylin and eosin-stained sections (left) with histologic tumour cell differentiation status in *Il33^{+/+}* and *Il33^{-/-}* PDAC mice (right). **c**, Trichrome staining in tumours from *Il33^{+/+}* and *Il33^{-/-}* PDAC mice ($n = 3$ per group). **d**, Immunohistochemistry for smooth muscle actin in tumours from *Il33^{+/+}* and *Il33^{-/-}* PDAC mice ($n = 3$ per group). **e**, Intratumoral ST2 expression on KPC cells in *Il33^{+/+}* and *Il33^{-/-}* orthotopic PDAC mice. **f**, ST2

expression on live KPC cells following rIL33 treatment in vitro (DRAQ7 stains dead cells) ($n = 3$ per group). **g**, KPC cell number, viability, proliferation (Ki-67), and apoptosis (annexin) following rIL33 treatment in vitro ($n = 3-6$ per group). Horizontal bars show median. n in **a-e** indicates individual mice analysed separately in at least two independent experiments with $n \geq 3$ per group. n in **f, g** indicates technical replicates and is representative of at least two independent experiments. P value determined by two-tailed Mann-Whitney test (**a**).



Extended Data Fig. 5 | ILC2s prime antigen-specific CD8⁺ T cells. **a**, Gating and frequency of TILC2s in ILC2-intact (diphtheria toxin (DT)-treated *Cd4^{Cre/+}Icos^{+/-}*) and ILC2-depleted (DT-treated *Cd4^{Cre/+}Icos^{fl-DTR/+}*) mice. **b**, Gating and frequency of OVA-specific CD8⁺ T cells in spleens from ILC2-intact and ILC2-depleted mice. OVA-specific T cells were detected as SIINFELK-tetramer⁺ cells. **c**, Gating and frequency of T_{CM} cells (CD45⁺CD3⁺CD8⁺CD44⁺CD62L⁺) in tumour, DLNs, and spleens from ILC2-intact and ILC2-depleted mice. **d**, ST2 expression on

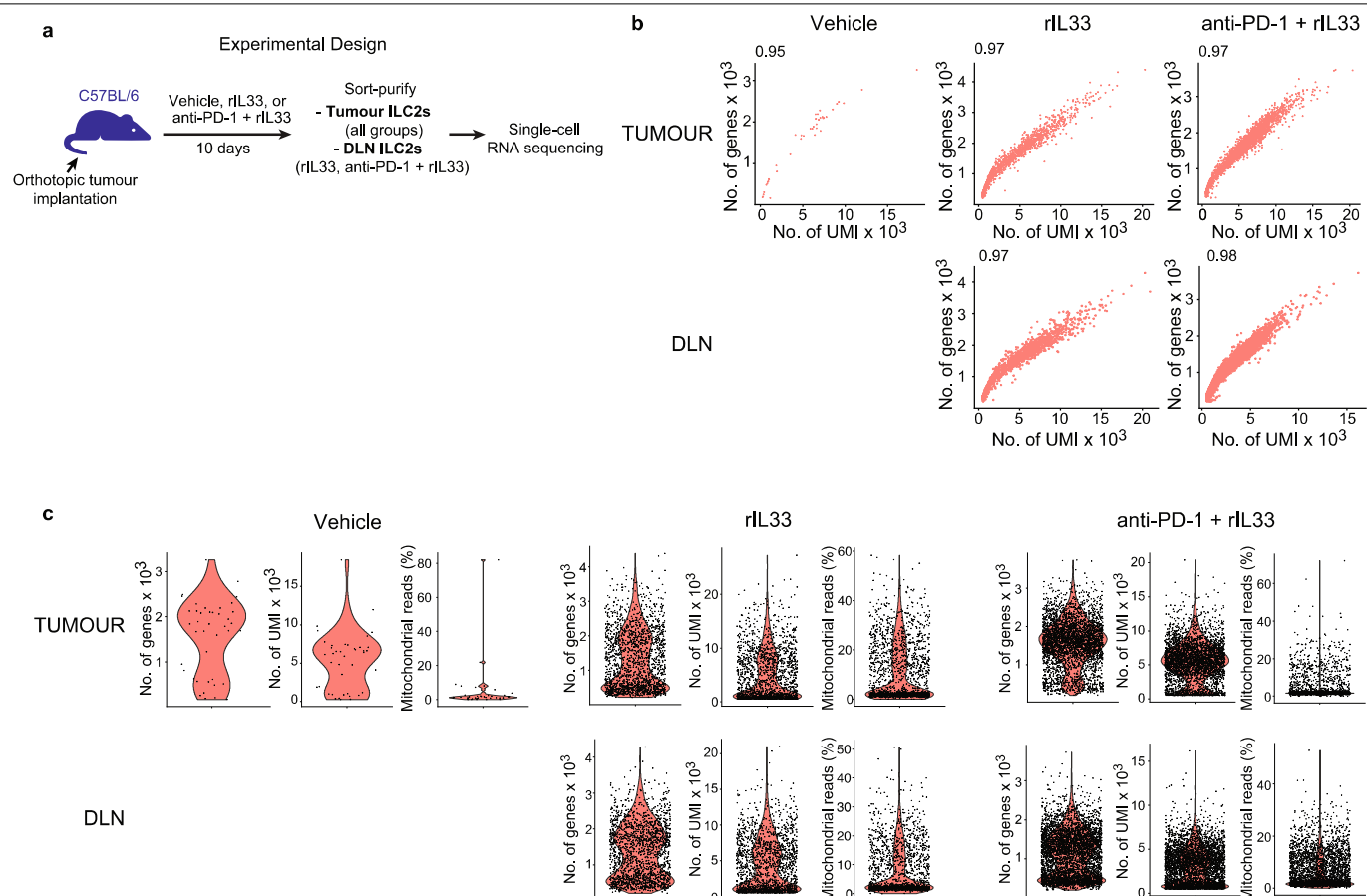
CD45⁺CD3⁺CD8⁺T cells after tumour implantation in PDAC mice. Data were collected 14 days after tumour implantation or at the time points indicated. Median \pm s.e.m; horizontal bars show median. *n* indicates individual mice analysed separately in at least two independent experiments with *n* \geq 2 per group. *P* values determined by two-tailed Mann-Whitney test (**a–c**) and two-way ANOVA with Tukey's multiple comparison post test (**d**, indicating comparison of tumour ILCs to all other groups).



Extended Data Fig. 6 | Immunophenotyping in rIL33-treated PDAC mice.

a, Tumour establishment of orthotopic and subcutaneous KPC-OVA PDAC tumours in vehicle (veh) and rIL33-treated mice. **b**, Gating (left) and frequency (right) of IL18R1 expression on tumour ILCs in subcutaneous (SQ) and orthotopic PDAC mice. **c**, Gating (left) and frequency (right) of splenic ILC2s following rIL33 treatment in orthotopic PDAC mice. **d**, Gating (left), frequency (middle), and number (right) of ILC2s following rIL33 treatment in subcutaneous PDAC mice. **e**, Gating (left) and frequency (right) of cytokine and PD-1 expression on tumour CD8⁺ T cells following rIL33 treatment in orthotopic PDAC mice. **f**, Frequency of immune cells in vehicle- and rIL33-treated

orthotopic PDAC mice. **g**, Gating strategy for identification of CD103⁺ DCs. **h**, Gating (left; tumours) and frequency (right) of ILC2s in tumours and DLNs from wild-type (WT) or ILC2-deficient (*Rora*^{fl/y} *IL7*^{Cre}) PDAC mice following rIL33 treatment. **i**, Gating (left) and frequency (right) of PD-1⁺ CD8⁺ T cells in tumours from rIL33-treated wild-type (WT) and *Batf3*^{-/-} mice. Data were collected 6 (**a**), 4 (**b**), 2 (**c**, **e**, **f**), 5 (**d**), 7 (**h**), and 3 (**i**) weeks after tumour implantation. Horizontal bars show median. n indicates individual mice analysed separately in at least two independent experiments with $n \geq 2$ per group. P values determined by χ^2 test (**a**), two-tailed Mann-Whitney test (all else).

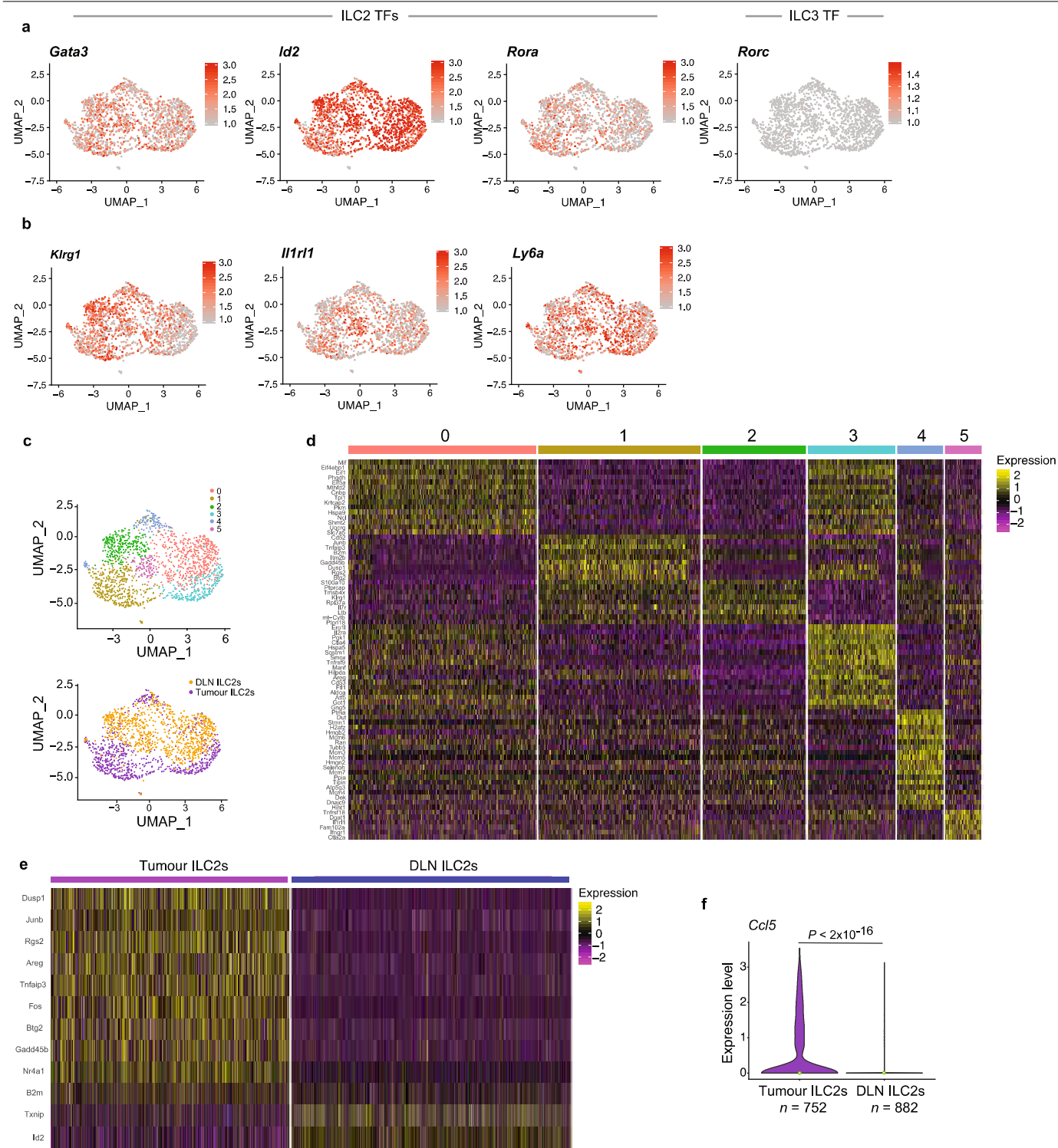


Extended Data Fig. 7 | scRNA-seq of tumour and DLN ILC2s in PDAC mice.

a, Experimental design for in vivo treatment, purification, and single-cell analysis of ILC2s. **b, c**, Quality metrics. **b**, Scatter plots showing, for each cell, the relationship between the number of UMIs and the number of genes.

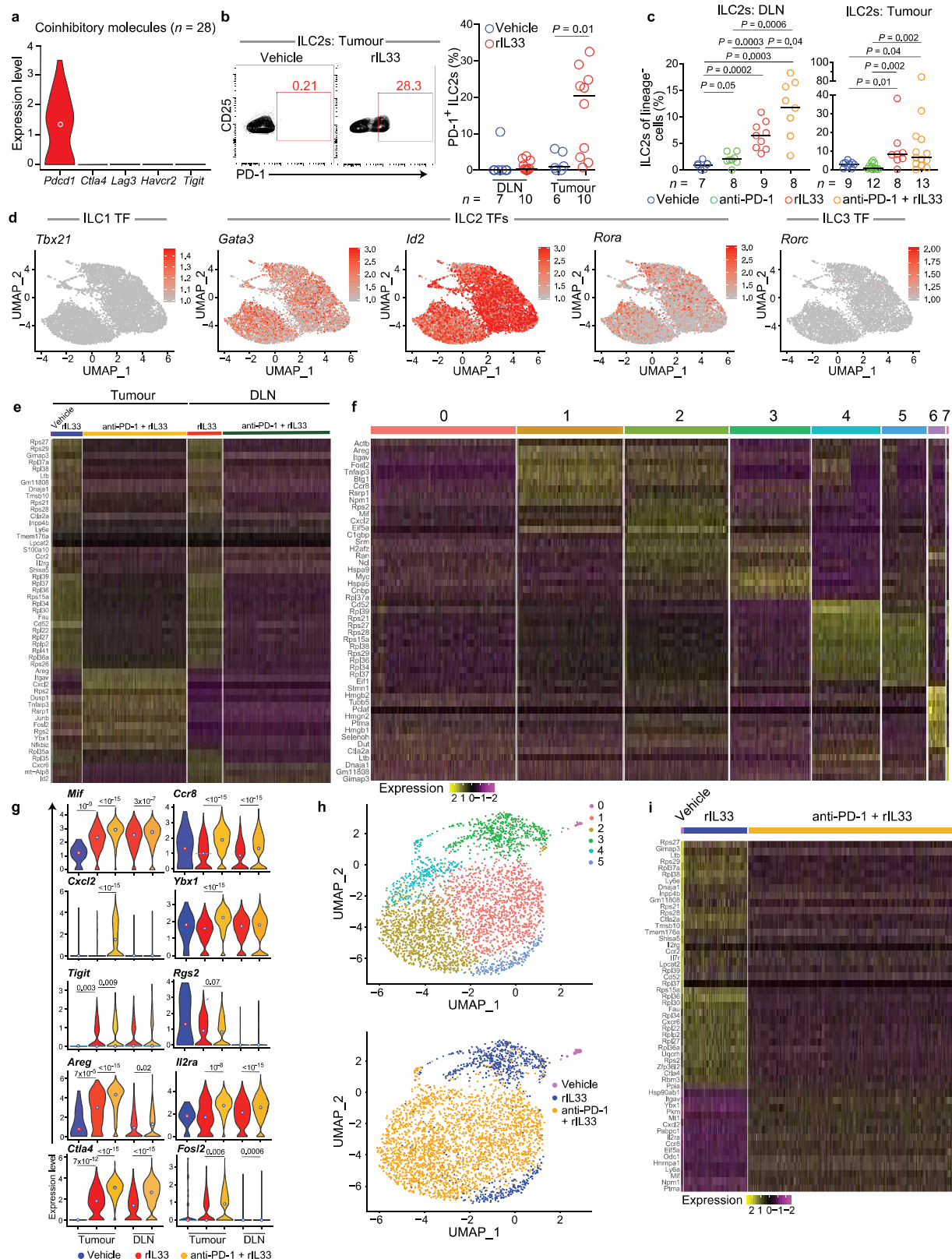
c, Violin plots showing the distribution of the number of genes (left), number of

UMIs (middle), and percentage of normalized reads from mitochondrial genes (right) in each treatment group (columns), and each tissue (rows). Each dot represents a single cell. For each treatment group and tissue, data represent pooled purified single cells from biological replicates of $n = 10$ (vehicle), $n = 5$ (rIL33), and $n = 5$ (anti-PD-1 + rIL33) PDAC mice.



Extended Data Fig. 8 | Activated ILC2s from tumours and DLNs have distinct transcriptional features. **a–d**, Single-cell analysis of 1,634 rIL33-activated tumour and DLN ILC2s (experimental design as in Extended Data Fig. 7a). UMAP plots show single cells (dots) in a nonlinear representation of the top 15 principal components. Expression of ILC2 (*Gata3*, *Id2*, *Rora*) and ILC3 (gene, *Rorc*; protein, RORγT) transcription factors (TFs) (**a**), ILC2 surface markers (**b**), and ILC clusters and tissues (tumour and DLN) (**c**). Expression of the ILC-1

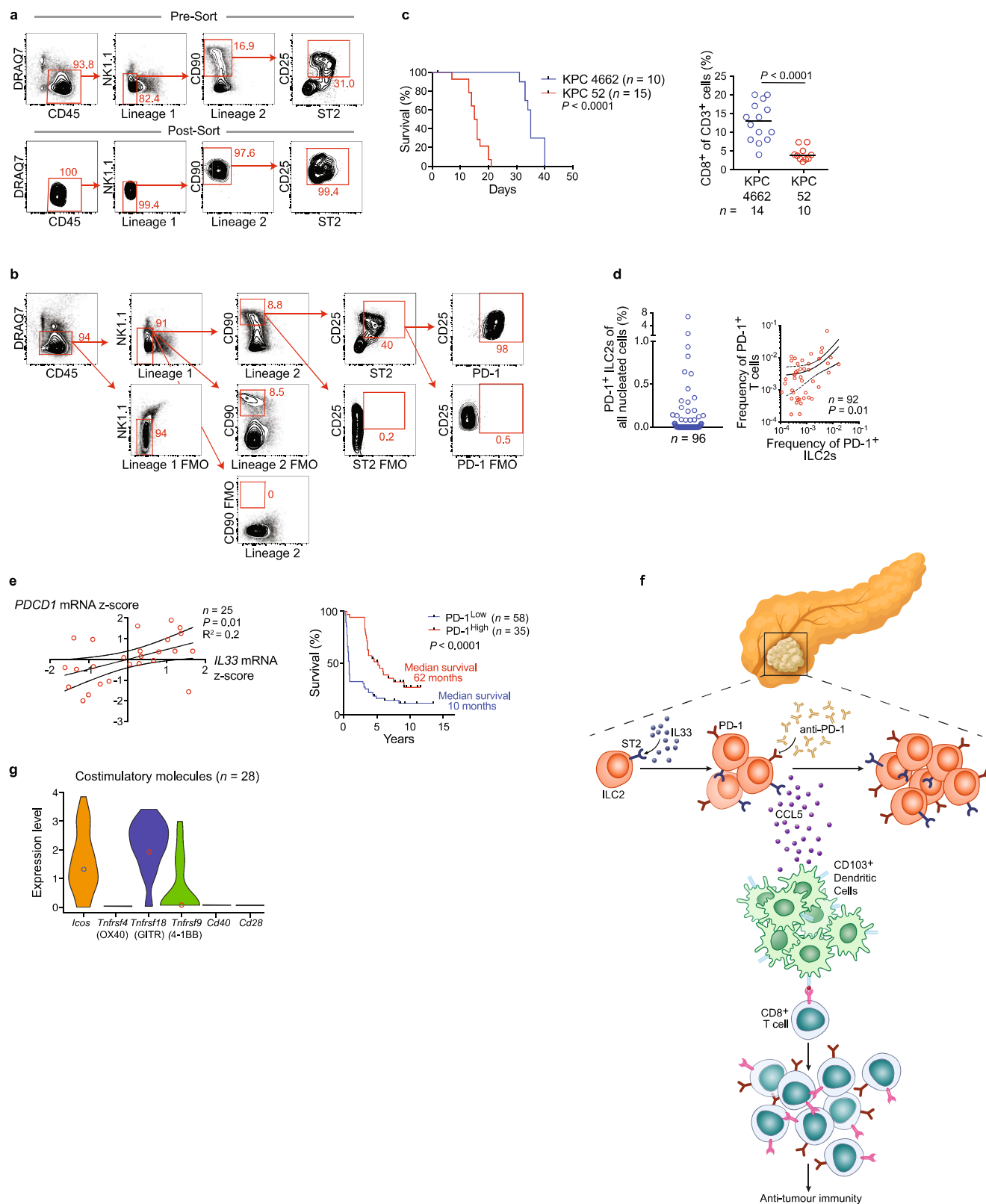
transcription factor *Tbx21* (T-bet) was undetectable. **d, e**, Differentially expressed genes by cluster (**d**) and tissue (**e**). **f**, Distribution of *Ccl5* expression from ILC2s in tumours and DLNs; violin plots show distribution with minima, maxima, and circle indicating median. Each dot in **a–c** represents a single cell. For each treatment group and tissue, data represent pooled purified single cells from biological replicates of $n = 5$ rIL33-treated PDAC mice. P value in **f** by two-sided pairwise Wilcoxon rank sum test.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Combined anti-PD-1 and rIL33 treatment induces a unique transcriptional profile in TILC2s. **a**, Expression of coinhibitory immune checkpoints in TILC2s in vehicle-treated PDAC mice by scRNA-seq. **b**, Gating and frequency of PD-1⁺ ILC2s in vehicle- and rIL33-treated PDAC mice. **c**, ILC2 frequency in treated PDAC mice. Corresponding tumour volumes, cell number, and scRNA-seq are shown in Fig. 4a–c. **d**, scRNA-seq of ILC2s from treated PDAC mice. Expression of ILC1 (gene, *Tbx21*; protein, T-bet), ILC2 (*Gata3*, *Id2*, *Rora*), and ILC3 (gene, *Rorc*; protein, Rorγt) transcription factors (TFs) in purified tumour and DLN ILC2s. Corresponding UMAP plots by cluster and treatment are depicted in Fig. 4c. **e–g**, Top differentially expressed genes by treatment and tissue (**e**), cluster (**f**), and distribution of expression for select

differentially expressed genes by treatment and tissue (**g**). **h**, UMAP plots of 3,415 single TILC2s in a nonlinear representation of the top 15 principal components. **i**, Differentially expressed genes in TILC2s by treatment. Each dot in **d**, **h** represents a single cell; in **d–i**, for each treatment group and tissue, data represent pooled purified single cells from biological replicates of $n=10$ (vehicle), $n=5$ (rIL33), and $n=5$ (anti-PD-1 + rIL33) PDAC mice (number of single cells for tumour: vehicle $n=28$, rIL33 $n=752$, rIL33 + anti-PD-1 $n=2,635$; for DLN: rIL33 $n=882$, rIL33 + anti-PD-1 $n=2,725$). Violin plots show distribution with minima, maxima, and circle indicating median. Horizontal bars show median. P values by two-tailed Mann–Whitney test (**b**, **c**) and two-sided pairwise Wilcoxon rank sum test (**g**).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Activated TILC2s express PD-1 and co-exist with PD-1⁺ T cells. **a, b,** Orthotopic PDAC mice (C57BL/6 WT, *Pdcd1*^{-/-}, CD45.1) were treated with 500 ng of carrier-free rIL33 daily for 10 days (experimental designs shown in Fig. 4e, f). Live, CD45⁺, lineage⁻, CD90⁺, CD25⁺, ST2⁺ TILC2s were sort-purified to 98% purity at day 10 post-implantation. TILC2s (5 × 10⁵ cells) were immediately transferred to orthotopic PDAC tumour-bearing ILC2-deficient (*Rora*^{fl/fl}/*IL7*^{Cre/+}) CD45.2 mice on days 7 and 14 post-tumour implantation via i.p. injection. Control mice received equivalent volumes of PBS via i.p. injection. **a,** Representative plots for TILC2 sort-purification (top) and post-sort purity (bottom). **b,** Representative plots showing PD-1 expression on sort-purified TILC2s from wild-type and CD45.1 mice in the experimental designs outlined in Fig. 4e, f. **c,** Survival and intratumoral CD8⁺ T cell frequency of orthotopic KPC 4662 and KPC 52 PDAC tumours; horizontal bars in **c** show median. **d,** Frequency of PD-1⁺ ILC2s (left) and correlation with PD-1⁺ T cells (right) in

human PDACs. **e,** Linear regression analysis of *IL33* and *PDCD1* mRNA in bulk tumour transcriptomes from short- and long-term human PDAC survivors (left) and survival association of PD-1⁺ cells in tumour tissue microarrays of short- and long-term PDAC survivors (right); high and low defined as higher or lower than the median for the cohort. **f,** Model linking the IL33–TILC2 axis to T cell immunity in PDAC. ©2019, Memorial Sloan Kettering Cancer Center. **g,** Distribution of expression of costimulatory molecules in untreated TILC2s by scRNA-seq. Experimental design as shown in Extended Data Fig. 7a; data represent pooled purified single cells from biological replicates of *n* = 10 (vehicle). Data are representative of purity and PD-1 expression on sorted TILC2s in two independent experiments with *n* ≥ 4 per group (**a, b**). *n* and data points denote individual mice and patients analysed separately. *P* values determined by two-tailed Mann–Whitney test (**c**), two-sided logrank test (**c, e**, survival curves) and linear regression (**d, e**).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Flow cytometric data were collected using FACSDiva (BD Biosciences, version 8.0.1). Pathologic slides were digitized using Panoramic Flash 250 (3Dhistech, Budapest Hungary) using Zeiss 20x/0.8NA objective and custom filters for A488, A546, A594 and A647.

Data analysis

Transcriptomic analyses was done using the Affymetrix Transcription Analysis Console (TAC) Software (version 3.0, Applied Biosystems, SST-RMA algorithm) to summarize the signal from array probesets (Ref 14). Flow cytometry data were analyzed on FlowJo (Treestar, versions 9.9.6, and 10.4.2). Digital cellular quantification was performed using custom macro written in Fiji/ImageJ (version 1.52n). For RNA sequencing, the expression dataset was loaded into Gene Set Enrichment Analysis (GSEA, version 3.0) to identify biological processes that were differentially expressed in experimental groups. Gene sets databases for antigen presentation and T cell mediated immunity were selected from MSIGDB (version 6.1). Single cell RNA sequencing analysis was performed with Cell Ranger Single Cell Software suite v3.0.2 (10x Genomics) and Seurat R package (version 3.1). Illumina Real Time Analysis Software (1.17) was used for RNA sequencing data quality assessment and data preparation. Statistical analysis was performed using Prism (GraphPad software, version 7.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw source data for all experiments included in this study are provided. Bulk RNA sequencing data is available under Gene Expression Omnibus (GEO) accession

number GSE129388. Single cell RNA sequencing data is available under GEO accession number GSE136720. Code for immune quantification is provided in Supplementary Data. All other data are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were determined based on our and other investigators experience with the respective cell lines used. No statistical methods were used as we observed many statistically significant effects in the data with the above methods of sample size selection without a priori sample size calculations.
Data exclusions	No data were excluded from the analyses.
Replication	All findings reported were reproducible and data shown are pooled from >=2 independent experiments, with comparable results in each experiment.
Randomization	6- to 12-week old mice were matched by age and sex and randomly assigned to specific treatment groups.
Blinding	No blinding was performed in experimental mouse interventions as knowledge of the treatment groups was required. To account for heterogeneity in tissue samples in human tumor cellular quantification, cellular frequencies for each tumor was calculated as the mean of 3 independent measurements of 3 randomly sampled biopsies of the same tumor sample. All histopathologic assessments were performed by a dedicated pancreatic pathologist who was blinded to the treatment groups. All digital quantification was automated, with all experimental groups quantified in an identical fashion.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Flow cytometry:
 Human (Name Fluorochrome Clone Supplier Catalog Lot Number Dilution)
 CD11b APC ICRF44 Biolegend 301310 B278347 1:20
 CD11c APC 3.9 ThermoFisher Scientific 17-0116-42 4329675 1:20
 CD127 FITC RDR5 ThermoFisher Scientific 11-1278-42 1971614 1:20
 CD16 APC CB16 ThermoFisher Scientific 17-0168-42 2013794 1:20
 CD19 AF700 HIB19 ThermoFisher Scientific 56-0199-42 2031137 1:20
 CD25 PerCP-Cy5.5 BC96 Biolegend 302626 B267552 1:20
 CD3 Alexa Fluor 700 OKT3 ThermoFisher Scientific 56-0037-42 1984155 1:20
 CD45 Pacific Blue HI30 Biolegend 304029 B256106 1:20
 CD5 APC L17F12 ThermoFisher Scientific 17-0058-42 1956856 1:20
 CD56 BV650 NCAM16.2 Biolegend 318344 B282104 1:20
 CRTH2 PE-Cy7 BM16 Biolegend 350118 B287080 1:20
 FcεR1 APC AER-37 ThermoFisher Scientific 17-5899-42 4300320 1:20
 GATA3 BV711 L50-823 BD Biosciences 565449 6140701 1:20
 ST2 PE hIL33Rcap ThermoFisher Scientific 12-9338-42 2041179 1:20
 TCRα/b APC IP26 Biolegend 306730 B251913 1:20

Mouse (Name Fluorochrome Clone Supplier Catalog Lot Number Dilution)

CD103 BV711 2E7 Biolegend 121435 B266330 1:40

CD11b APC M1/70 ThermoFisher Scientific 45-0112-82 1929457 1:80

CD11c APC HL3 BD Biosciences 550261 9129856 1:200

CD127 FITC A7R34 ThermoFisher Scientific 11-1271-85 2083447 1:50

CD19 Alexa Fluor 700 1D3 ThermoFisher Scientific 56-0193-82 4345832 1:80

CD25 PerCP-Cy5.5 PC61.5 ThermoFisher Scientific 45-0251-82 4289647 1:80

CD3 Alexa Fluor 700 17A2 ThermoFisher Scientific 56-0032-82 4336536 1:80

CD4 BV786 RM4-5 BD Biosciences 563727 7166502 1:200

CD44 PerCP-Cy5.5 IM7 ThermoFisher Scientific 45-0441-82 1984139 1:80

CD45 Pacific Blue 30-F11 Biolegend 103126 B253970 1:200

CD45.1 BV711 A20 Biolegend 110739 B245194 1:40

CD5 APC 53-7.3 BD Biosciences 550035 8187632 1:200

CD62L APC MEL-14 BD Biosciences 561919 7047558 1:80

CD8 Alexa Fluor 700 53-6.7 ThermoFisher Scientific 56-0081-82 4329739 1:160

CD90.2 BV786 53-2.1 BD Biosciences 564365 8186556 1:100

F4/80 PE-Cy5 BM8 ThermoFisher Scientific 15-4801-82 4316699 1:80

FceR1 APC MAR-1 ThermoFisher Scientific 17-5898-82 2095487 1:160

FoxP3 APC FJK-16S ThermoFisher Scientific 17-5773-82 4313491 1:20

Gata3 PE-Cy7 L50-823 BD Biosciences 560405 8242846 1:5

Gr-1 BV605 RB6-8C5 Biolegend 108439 B219337 1:40

IFN-g APC-Cy7 XMG1.2 BD Biosciences 561479 7153814 1:100

IL4 BV650 11B11 BD Biosciences 564004 7256960 1:100

IL5 PE TRFK5 ThermoFisher Scientific 12-7052-82 4312191 1:80

Ly6C PerCP-Cy5.5 AL-21 BD Biosciences 560525

Ly6G AF700 1A8 BD Biosciences 561236 7200759 1:100

MHC-II Alexa Fluor 700 M5/114.15.2 ThermoFisher Scientific 56-5321-82 1919519 1:330

NK1.1 BV650 PK136 BD Biosciences 564143 8162883 1:80

PD1 BV605 J43 BD Biosciences 563059 7047607 1:100

Ror-gt BV786 Q31-378 BD Biosciences 564723 7117673 1:200

ST2 PE-Cy7 RMST2-2 ThermoFisher Scientific 25-9335-82 2035263 1:80

Tbet BV711 Q4-46 BD Biosciences 563320 8150719 1:20

TNF-a BV510 MP6-XT22 BD Biosciences 563386 8138693 1:80

SINFEKL tetramer MBL international N/A TB-5001-1 T1708003 1:50

Other experiments:

Anti-alpha smooth muscle Actin Antibody (PE-Cy7) Abcore Inc 1A4 AC12-0159-17 Lot # not available 1:500 - 1:1000

Biotinylated Goat Anti-Rabbit IgG Vector Laboratories Polyclonal BA-1000 Lot # not available 1:200

Biotinylated Horse Anti-Goat IgG Vector Laboratories Polyclonal BA-9500 Lot # not available 1:200

Biotinylated Rabbit Anti-Goat IgG Vector Laboratories Polyclonal BA-5000 Lot # not available 1:200

CD45 Antibody (2B11 + PD7/26) DAKO 2B11 + PD7/26 NBP2-34287 Lot # not available 1:200 - 1:400

GATA3 (L50-823) Mouse Monoclonal Antibody Cell Marque L50-823 390M 0000032957 1:400

HIL-33 PE MAB R&D Systems/Fisher Scientific Polyclonal AF3625 ACAP0216021 1:13 - 1:40

IL-33R (ST2) Monoclonal Antibody (RMST2-2), PE-Cyanine7, eBioscience ThermoFisher Scientific RMST2 25-9335-82 4298148 1:80

InVivoMAb anti-mouse CD16/CD32 BioXCell 2.4G2 BE0307 636517D1 1:20

InVivoMAb anti-mouse Thy1.2 (CD90.2) BioXCell 30-H12 BE0066 Lot # not available 500 ug/mouse

InVivoPlus anti-mouse CD4 (11723081) BioXCell GK1.5 BP0003 628316D1B 500ug/mouse

InVivoPlus anti-mouse CD8a (11723145) BioXcell 2.43 BP0061 653047M2B 500 ug/mouse

InVivoPlus rat IgG2b I sotype control (11723013) BioXcell LTF-2 BP0090 629817O1 500ug/mouse

Ki-67 Monoclonal Antibody (SolA15) ThermoFisher Scientific SolA15 46-5698-80 4296883 1:333

Microglia Marker - Iba1 Antibody Wako Polyclonal 019-19741 Lot # not available 1:200

Mouse IL-33 Antibody R&D Systems Polyclonal AF3626 PDJ1519052 1:400

PD-1 (NAT105) Mouse Monoclonal Antibody Ventana Roche NAT105 760-4895 N/A 1:100

Recombinant Anti-CD11b antibody [EPR1344] Abcam EPR1344 AB133357 Lot # not available 1:1000

Recombinant Anti-Cytokeratin 19 antibody [EP1580Y] - Cytoskeleton Marker Abcam EP1580Y AB52625 GR249900 1:200 - 1:500

CD3, Polyclonal, Unconjugated, Affinity isolated Antibody DAKO F7.2.38 A0452 Lot # not available 1:400

Validation

All antibodies were validated by the manufacturer and used per their instructions. In our experiments, isotype and/or FMO control samples were included. Specific assessments of differential expression in our model system was performed by comparing expression characteristics in tumor tissue to non-tumor adjacent normal tissue (adjacent normal pancreas), tumor regional lymphoid organs (draining lymph node), and non-tumor lymphoid organs (spleen).

All antibodies were validated by the manufacturer and used per their instructions. In our experiments, isotype and/or FMO control samples were included. Specific assessments of differential expression in our model system was performed by comparing expression characteristics in tumor tissue to non-tumor adjacent normal tissue (adjacent normal pancreas), tumor regional lymphoid organs (draining lymph node), and non-tumor lymphoid organs (spleen).

Additional information on validation can be found on the manufacturers' websites listed below:

Mouse flow cytometry antibodies:

BD Biosciences, CD11c APC HL3, # 550261 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-mouse-antibodies/cell-surface-antigens/apc-hamster-anti-mouse-cd11c-hl3/p/550261>

BD Biosciences, NK1.1 BV650 PK136, # 564143

<https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-mouse-antibodies/cell-surface-antigens/bv650-mouse-anti-mouse-nk-11-pk136/p/564143>

BD Biosciences, CD4 BV786 RM4-5, # 563727 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/surface-markers/mouse/bv786-rat-anti-mouse-cd4-rm4-5/p/563727>

BD Biosciences, CD62L APC MEL-14, # 561919 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/regulatory-t-cells/surface-markers/mouse/apc-rat-anti-mouse-cd62l-mel-14/p/561919>

BD Biosciences, CD19 AF700 1D3, # 562956
<https://www.bdbiosciences.com/us/applications/research/stem-cell-research/hematopoietic-stem-cell-markers/mouse/negative-markers/alexa-fluor-700-rat-anti-mouse-cd19-1d3/p/557958>

BD Biosciences, Ly6C PerCP-Cy5.5 AL-21, # 560525 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-mouse-antibodies/cell-surface-antigens/percp-cy55-rat-anti-mouse-ly-6c-al-21/p/560525>

BD Biosciences, Ly6G AF700 1A8, # 561236 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-mouse-antibodies/cell-surface-antigens/alexa-fluor-700-rat-anti-mouse-ly-6g-1a8/p/561236>

BD Biosciences, PD1 BV605 J43, # 563059 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/regulatory-t-cells/surface-markers/mouse/bv605-hamster-anti-mouse-cd279-j43/p/563059>

BD Biosciences, TNF-α BV510 MP6-XT22, # 563386 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/intracellular-markers/cytokines-and-chemokines/mouse/bv510-rat-anti-mouse-tnf-mp6-xt22/p/563386>

BD Biosciences, IFN-γ APC-Cy7 XMG1.2, # 561479 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/intracellular-markers/cytokines-and-chemokines/mouse/apc-cy7-rat-anti-mouse-ifn-xmg12/p/561479>

BD Biosciences, CD90.2 BV786 53-2.1, # 564365 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-mouse-antibodies/cell-surface-antigens/bv786-rat-anti-mouse-cd902-53-21/p/564365>

BD Biosciences, Tbet BV711 Q4-46, # 563320 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/intracellular-markers/cell-signalling-and-transcription-factors/mouse/bv711-mouse-anti-tbet-q4-46/p/563320>

BD Biosciences, Rorγt BV786 Q31-378, # 564723 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/cell-biology-reagents/cell-biology-antibodies/bv786-mouse-anti-mouse-rort-q31-378/p/564723>

BD Biosciences, Gata3 PE-Cy7 L50-823, # 560405 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-2-cells/intracellular-markers/cell-signalling-and-transcription-factors/human/pe-cy7-mouse-anti-gata3-l50-823/p/560405>

BD Biosciences, IL4 BV650 11B11, # 564004 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-mouse-antibodies/intracellular-antigens/bv650-rat-anti-mouse-il-4-11b11/p/564004>

BD Biosciences, CD5 APC 53-7.3, # 550035 <https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/regulatory-t-cells/surface-markers/mouse/apc-rat-anti-mouse-cd5-53-73/p/550035>

Biolegend, CD45 Pacific Blue 30-F11, # 103126 <https://www.biolegend.com/en-us/products/pacific-blue-anti-mouse-cd45-antibody-3102>

Biolegend, CD45.1 BV711 A20, # 110739 <https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-cd45-1-antibody-8925>

Biolegend, Gr-1 BV605 RB6-8C5, # 108439 <https://www.biolegend.com/en-us/products/brilliant-violet-605-anti-mouse-ly-6g-ly-6c-gr-1-antibody-8724>

Biolegend, CD103 BV711 2E7, # 121435 <https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-cd103-antibody-14411>

ThermoFisher Scientific, CD3 Alexa Fluor 700 17A2, # 56-0032-82 <https://www.thermofisher.com/antibody/product/CD3-Antibody-clone-17A2-Monoclonal/56-0032-82>

ThermoFisher Scientific, CD19 Alexa Fluor 700 1D3, # 56-0193-82 <https://www.thermofisher.com/antibody/product/CD19-Antibody-clone-eBio1D3-1D3-Monoclonal/56-0193-82>

ThermoFisher Scientific, FcεR1 APC MAR-1, # 17-5898-82 <https://www.thermofisher.com/antibody/product/FcεR1-alpha-Antibody-clone-MAR-1-Monoclonal/17-5898-82>

ThermoFisher Scientific, CD8 Alexa Fluor 700 53-6.7, # 56-0081-82 <https://www.thermofisher.com/antibody/product/CD8a-Antibody-clone-53-6-7-Monoclonal/56-0081-82>

ThermoFisher Scientific, CD11b APC M1/70, # 45-0112-82 <https://www.thermofisher.com/antibody/product/CD11b-Antibody-clone-M1-70-Monoclonal/45-0112-82>

ThermoFisher Scientific, F4/80 PE-Cy5 BM8, # 15-4801-82 <https://www.thermofisher.com/antibody/product/F4-80-Antibody-clone-BM8-Monoclonal/15-4801-82>

ThermoFisher Scientific , MHC-II Alexa Fluor 700 M5/114.15.2, # 56-5321-82 <https://www.thermofisher.com/antibody/product/MHC-Class-II-I-A-I-E-Antibody-clone-M5-114-15-2-Monoclonal/56-5321-82>

ThermoFisher Scientific CD44 PerCP-Cy5.5 IM7, # 45-0441-82 <https://www.thermofisher.com/antibody/product/CD44-Antibody-clone-IM7-Monoclonal/45-0441-82>

ThermoFisher Scientific CD127 FITC A7R34, # 11-1271-85 <https://www.thermofisher.com/antibody/product/CD127-Antibody-clone-A7R34-Monoclonal/11-1271-85>

ThermoFisher Scientific CD25 PerCP-Cy5.5 PC61.5, # 45-0251-82 <https://www.thermofisher.com/antibody/product/CD25-Antibody-clone-PC61-5-Monoclonal/45-0251-82>

ThermoFisher Scientific IL5 PE TRFK5, # 12-7052-82 <https://www.thermofisher.com/antibody/product/IL-5-Antibody-clone-TRFK5-Monoclonal/12-7052-82>

ThermoFisher Scientific ST2 PE-Cy7 RMST2-2, # 25-9335-82 <https://www.thermofisher.com/antibody/product/IL-33R-ST2-Antibody-clone-RMST2-2-Monoclonal/25-9335-82>

ThermoFisher Scientific FoxP3 APC FJK-16S, # 17-5773-82 <https://www.thermofisher.com/antibody/product/FOXP3-Antibody-clone-FJK-16s-Monoclonal/17-5773-82>

Human flow cytometry antibodies:

BD Biosciences, GATA3 BV711 L50-823, # 565449 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/cell-biology-reagents/cell-biology-antibodies/bv711-mouse-anti-gata3-l50-823/p/565449>

Biolegend, CD56 BV650 NCAM16.2, # 318344 <https://www.biolegend.com/en-us/products/brilliant-violet-650-anti-human-cd56-ncam-antibody-8780>

Biolegend, CD25 PerCP-Cy5.5 BC96, # 302626 <https://www.biolegend.com/en-us/products/percpcyanine55-anti-human-cd25-antibody-4231>

Biolegend, CD11b APC ICRF44, # 301310 <https://www.biolegend.com/en-us/products/apc-anti-human-cd11b-antibody-765>

Biolegend, CRTH2 PE-Cy7 BM16, # 350118 <https://www.biolegend.com/en-us/products/pe-cy7-anti-human-cd294-crth2-antibody-8815>

Biolegend, CD45 Pacific Blue HI30, # 304029 <https://www.biolegend.com/en-us/products/pacific-blue-anti-human-cd45-antibody-3331>

Biolegend, TCRa/b APC IP26, # 306730 <https://www.biolegend.com/en-us/products/alexa-fluor-700-anti-human-tcr-alpha-beta-antibody-12517>

ThermoFisher Scientific, CD3 Alexa Fluor 700 OKT3, # 56-0037-42 <https://www.thermofisher.com/antibody/product/CD3-Antibody-clone-OKT3-Monoclonal/56-0037-42>

ThermoFisher Scientific, CD127 FITC RDR5, # 11-1278-42 <https://www.thermofisher.com/antibody/product/CD127-Antibody-clone-eBioRDR5-Monoclonal/11-1278-42>

ThermoFisher Scientific, CD11c APC 3.9, # 17-0116-42 <https://www.thermofisher.com/antibody/product/CD11c-Antibody-clone-3-9-Monoclonal/17-0116-42>

ThermoFisher Scientific, CD16 APC CB16, # 17-0168-42 <https://www.thermofisher.com/antibody/product/CD16-Antibody-clone-eBioCB16-CB16-Monoclonal/17-0168-42>

ThermoFisher Scientific, ST2 PE hIL33Rcap, # 12-9338-42 <https://www.thermofisher.com/antibody/product/IL-33R-ST2-Antibody-clone-hIL33Rcap-Monoclonal/12-9338-42>

ThermoFisher Scientific, CD5 APC L17F12, # 17-0058-42 <https://www.thermofisher.com/antibody/product/CD5-Antibody-clone-L17F12-Monoclonal/17-0058-42>

ThermoFisher Scientific, CD19 AF700 HIB19, # 56-0199-42 <https://www.thermofisher.com/antibody/product/CD19-Antibody-clone-HIB19-Monoclonal/56-0199-42>

ThermoFisher Scientific, FcεR1 APC AER-37, # 17-5899-42 <https://www.thermofisher.com/antibody/product/FcεR1-alpha-Antibody-clone-AER-37-CRA1-Monoclonal/17-5899-42>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

All tumor cell lines were derived from KPC mice that were backcrossed more than 10 generations with C57Bl/6 mice. KPC

Cell line source(s)	4662 and KPC 4662 OVA (V6 clone) cells were derived from Pdx1-Cre;LSL-KrasG12D/+;LSL-Trp53R172H/+ (Ref 37). KPC 8-1, 18-3, and 52 cells derived from Ptf1a-Cre;LSL-KrasG12D/+;LSL-Trp53R172H/+ mice.
Authentication	All cell lines were authenticated as bonafide pancreatic cancer cell lines. This was based on histopathologic verification by a dedicated pancreatic cancer pathologist that these cell lines generate tumors on intra-pancreatic implantation that faithfully recapitulate features of both human pancreatic cancers and pancreatic cancers that develop in spontaneous genetically engineered mice.
Mycoplasma contamination	Cell lines were regularly tested using MycoAlert Mycoplasma Detection Kit (Lonza). None of the cell lines used in this study tested positive for Mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	C57BL/6 (wild type, WT, CD45.2), C57BL/6 CD45.1, Pdc1-/-, Rag2-/-, Rag2-/-gc-/-, and Batf3-/-, mice were purchased from Jackson Labs. Il33-/-, Il33Cit/+ were a gift from M.J. Rosen, and have been previously described (Il33-/- PMID: 20937871; Il33Cit/+). Cd4Cre/+; Icosfl-Dtr/+ and Il7rCre/+; Rorafl/fl were a gift from A.N.J. McKenzie and have been previously described (Refs. 16). Pdx-Cre; LSL-Kras-G12D; LSL-Trp53R172H/+ (KPC) mice have been previously described (ref 4). For all experiments, 6- to 12-week old mice were matched by age and sex and randomly assigned to specific treatment groups, with at least two independent experiments performed throughout. Both male and female animals were utilized. KPC mice were sacrificed when tumors were detectable by ultrasound. Animals were bred and maintained in a specific pathogen-free animal facility at Memorial Sloan Kettering Cancer Center.
Wild animals	No wild animals were used.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	All animal studies were in accordance with the Institutional Animal Care and Use Committee protocol at Memorial Sloan Kettering Cancer Center.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	<p>All tumor samples were from patients with surgically resected primary pancreatic ductal adenocarcinomas. Patients in the tissue microarray cohort have been previously described (Ref. 3). Clinical characteristics of patients in the flow cytometry and transcriptomic cohorts are outlined in Supplementary Tables 1 and 2, respectively, and are provided below.</p> <p>Flow cytometry cohort (n=25); presented in Supplementary Table 1 Gender: Male 16 (64%); Female 9 (36%) Age, Median (Range): 65 (50-77) Tumor Location: Head 20 (80%), Body/Tail 5 (20%) Procedure: Distal Pancreatectomy 5 (20%), Pancreaticoduodenectomy 20 (80%) Pathological State: IA 3 (12%), IB 2 (8%), IIA 4 (16%), IIB 14 (56%), III 2 (8%), IV 0 (0%) pT: 1 3 (12%), 2 6 (24%), 3 16 (64%), 4 0 (0%) pN: 0 9 (36%), 1 14 (56%), 2 2 (8%) pM: 0 25 (100%), 1 0 (0%) Margin: Positive 5 (20%), Negative 20 (80%) Adjuvant Treatment: Yes 19 (76%), No 6 (24%) Neoadjuvant Treatment: Yes 11 (44%), No 14 (56%)</p> <p>Transcriptomic cohort (short-term survivors [n=12] vs. long-term survivors [n=13]); presented in Supplementary Table 2 Male: 6 (50%) vs. 5 (38%) Female: 6 (50%) vs. 8 (62%) Age, Median (Range): 76 (53-83) vs. 58 (45-88) Tumor Location: Head, 8 (67%) vs. 10 (77%) Body/Tail, 4 (33%) vs. 3 (23%) Procedure: Distal Pancreatectomy, 4 (33%) vs. 3 (23%) Pancreaticoduodenectomy, 8 (67%) vs. 10 (77%) Total Pancreatectomy, 0 (0%) vs. 0 (0%) Pathological Stage: I, 0 (0%) vs. 0 (0%) II, 9 (75%) vs. 12 (92%) III, 1 (8%) vs. 1 (8%) IV, 2 (17%) vs. 0 (0%) pT:</p>
----------------------------	--

1, 0 (0%) vs. 0 (0%)
 2, 0 (0%) vs. 0 (0%)
 3, 10 (83%) vs. 12 (92%)
 4, 2 (17%) vs. 1 (8%)
 pN:
 0, 5 (42%) vs. 6 (46%)
 1, 7 (58%) vs. 7 (54%)
 pM:
 0, 10 (83%) vs. 13 (100%)
 1, 2 (17%) vs. 0 (0%)
 Margin:
 Positive, 4 (33%) vs. 1 (8%)
 Negative, 8 (67%) vs. 12 (92%)
 Adjuvant Treatment:
 Yes, 9 (75%) vs. 10 (77%)
 No, 3 (25%) vs. 3 (23%)
 Unknown, 0 (0%) vs. 0 (0%)

Recruitment

All pancreatic ductal adenocarcinoma patients eligible for surgical resection at Memorial Sloan Kettering Cancer Center were recruited to participate in an Institutional Review Board-approved protocol. All patients who provided informed consent had samples collected; all study procedures were conducted in strict compliance with all ethical and institutional regulations. As patients were only recruited at Memorial Sloan Kettering Cancer Center, there is the potential for institution-specific selection bias. We do not believe that this potential bias would impact the results of this study.

Ethics oversight

All tissues were collected at Memorial Sloan Kettering Cancer Center under study protocol #15-149 and was approved by the Memorial Sloan Kettering Cancer Center Institutional Review Board. Informed consent was obtained for all patients. The study was in strict compliance with all institutional ethical regulations. All tumor samples were surgically resected primary pancreatic ductal adenocarcinomas.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Bone Marrow Harvest

Bone marrow was harvested from CD45.2 congenically-labeled donor mice, filtered through a 70mm filter, centrifuged, and resuspended in sterile PBS to a concentration of 1e8 live cells per 200ul. CD45.1 congenically-labeled C57BL/6J recipient mice were irradiated (5.5 Gy X2, 6 hours apart) 24 hours before bone marrow transplant and were maintained on endofloxacin water for 4 weeks post irradiation. Single cell suspensions of CD45.2 bone marrow chimera in sterile PBS (1e8 live cells per recipient mouse) was transplanted to each recipient mice by retroorbital injection. Reconstitution was confirmed by flow cytometry of the peripheral blood at 4 and 8 weeks post transplantation.

ILC2 adoptive transfer

CD45.1 C57BL/6 or Pdcd1^{-/-} orthotopic PDAC mice were treated with 500 ng of carrier-free recombinant murine IL33 (R&D Systems) in sterile PBS daily for 10 days. Live, CD45⁺, lineage⁻, CD90⁺, CD25⁺, ST2⁺ tumor ILC2s were sort-purified to 98% purity at day 10 post-implantation using an Aria Cell sorter (BD Biosciences). 5 x 10⁵ tumor ILC2s were immediately transferred to orthotopic PDAC tumor bearing Il7rCre/+Roralf/fl CD45.2 mice on days 7 and 14 post tumor implantation via i.p. injection. Control mice received equivalent volumes of PBS via i.p. injections. aPD-1 treatment in recipient mice was initiated on the day of ILC2 cell transfer.

Human tumor transcriptomic profiling

Patient subsets were randomly selected to undergo transcriptomic profiling as previously described (Ref 14). Patients in the TMA cohort with tumor tissue available for transcriptomic assessment were included in analyses in Figure 1b to allow protein confirmation of RNA expression. Extracted RNA was qualified on an Agilent BioAnalyzer and quantified by fluorometry (Ribogreen). Preparation of RNA for whole-transcriptome expression analysis was done using the WT Pico Reagent Kit (Affymetrix). Reverse transcription was initiated at the poly-A tail as well as throughout the entire length of RNA to capture both coding and multiple forms of non-coding RNA. RNA amplification was achieved using low-cycle PCR followed by linear amplification using T7 in vitro transcription technology. The cRNA was then converted to biotinylated sense-strand DNA hybridization targets. The prepared target was hybridized to GeneChip Human Transcriptome Array 2.0 (Affymetrix). Wash and scan was performed using the GeneChip Hybridization, Wash and Stain Kit using a Fluidics Station 450/250. Arrays were scanned using the GeneChip Scanner 3000.

Mouse RNA sequencing

Tissues from orthotopic PDAC mice (n=6) were harvested and dissociated into single-cell suspension as described above. Tumor-infiltrating leukocytes were positively selected by magnetically-activated cell sorting using mouse CD45 MicroBeads (Miltenyi Biotec). Purification of magnetically-activated sorted cells was confirmed by flow cytometry and was >95%. RNA was isolated from the sorted cells using an RNeasy Plus Mini Kit (Qiagen). Poly(A) capture and paired-end RNA sequencing were performed by the Memorial Sloan Kettering Integrated Genomics Core Facility. Specifically, after RiboGreen quantification and quality control by Agilent BioAnalyzer, 500 ng of total RNA underwent polyA selection and TruSeq library preparation according to instructions provided by Illumina (TruSeq Stranded mRNA LT Kit, catalog # RS-122-2102), with 8 cycles of PCR. Samples were barcoded and run on a HiSeq 4000 in a 100bp/100bp paired-end run, using the HiSeq 3000/4000 SBS Kit (Illumina). An average of 83 million paired reads was generated per sample. Ribosomal reads represented at most 0.03% of the total reads generated, and the percentage of mRNA bases averaged 76.6%.

Mouse single cell RNA sequencing

Single-cell suspensions of FACS-purified ILC2 cells from vehicle, IL33 alone and IL33+ PD1-treated pancreatic KPC tumors and mesenteric draining lymph nodes were prepared (purity >98%). Single-cell RNA-seq libraries were prepared according to 10x Genomics specifications (Chromium Single Cell V(D)J User Guide PN-1000006, 10x Genomics, Pleasanton, CA, USA). Four independent cellular suspensions (85-90% viable) at a concentration between 90-200 cells/ul, were loaded onto the 10x Genomics Chromium platform to generate Gel Beads-in-Emulsion (GEM), targeting about 2000 single cells per sample. After GEM generation, the samples were subjected to an incubation at 53°C for 45 min in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad, Hercules) to generate polyA cDNA barcoded at the 5' end by the addition of a template switch oligo (TSO) linked to a cell barcode and Unique Molecular Identifiers (UMIs). GEMs were broken and the single-strand cDNA was cleaned up with DynaBeads MyOne Silane Beads (Thermo Fisher Scientific, Waltham, MA). The cDNA was amplified for 16 cycles (98°C for 45 sec; 98°C for 20s, 67°C for 30s, 72°C for 1hr). Quality of the cDNA was assessed using an Agilent Bioanalyzer 2100 (Santa Clara, CA), obtaining a product of about 1200bp. 50ng of cDNA was enzymatically fragmented, end repaired, A-tailed, subjected to a double-sided size selection with SPRIselect beads (Beckman Coulter, Indianapolis, IN) and ligated to adaptors provided in the kit. A unique sample index for each library was introduced through 14 cycles of PCR amplification using the indexes provided in the kit (98°C for 45 s; 98°C for 20s, 54°C for 30s, and 72°C for 20 s x 14 cycles; 72°C for 1 min; held at 4°C). Indexed libraries were subjected a second double-sided size selection, and libraries were then quantified using Qubit fluorometric quantification (Thermo Fisher Scientific, Waltham, MA). The quality was assessed on an Agilent Bioanalyzer 2100, obtaining an average library size of 450bp. No treatment samples had concentrations below detectable limits, cDNA amplification was done with 18 cycles, and sample Index with 16 cycles. Libraries were diluted to 10nM and clustered using a NovaSeq600 on a pair end read flow cell and sequenced for 28 cycles on R1 (10x barcode and the UMIs), followed by 8 cycles of I7 Index (sample Index), and 89 bases on R2 (transcript), obtaining about 100M clusters per sample, except for tumors from vehicle treated mice which was clustered at about 10M. Primary processing of sequencing images was done using Illumina's Real Time Analysis software (RTA). 10x Genomics Cell Ranger Single Cell Software suite v3.0.2 (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger>) was used to perform sample demultiplexing, alignment to mouse genomic reference mm10, filtering, UMI counting, single-cell 5' end gene counting and performing quality control using the manufacturer parameters. Data from approximately 11, 000 single cells that passed quality control were obtained with approximately 41,000 mean reads per cell (48% sequencing saturation).

Flow Cytometry

Mouse and human PDAC tumors and adjacent pancreata were mechanically dissociated and incubated in collagenase (collagenase II for murine tumors, collagenase IV for human tumors, both 5 mg/ml; Worthington Biochemical Corp., Fisher Scientific), DNase I (0.5 mg/ml; Roche Diagnostics), and Hank's balanced salt solution (Gibco, Fisher Scientific) for 30 minutes at 37°C. Digestion was then quenched with fetal bovine serum (FBS, Life Technologies), and cells were filtered sequentially through 100- and 40-mm nylon cell strainers (Falcon, Fisher Scientific). Tumors, adjacent pancreata, and lymph nodes were then mechanically dissociated and filtered through 100- and 40-mm nylon cell strainers (Falcon, Fisher Scientific) using PBS with 1% FBS (Life Technologies). Spleens were mechanically dissociated and filtered through 70- and 40-mm nylon cell strainers (Falcon, Fisher Scientific) using PBS with 1% FBS, followed by RBC lysis (RBC lysis buffer, ThermoFisher Scientific). Mouse Fc receptors were blocked with FcεRIII/II-specific antibody (1 ug per 1 x 10⁶ cells; 2.4G2; Bio X Cell).

Instrument

Flow cytometry data was collected on a BD LSR Fortessa (BD Biosciences). Flow cytometry sorting was performed on a BD FACS Aria (BD Biosciences).

Software

Flow cytometry data was analyzed using FlowJo (Treestar versions 9.9.6, and 10.4.2)

Cell population abundance

Tumor-infiltrating leukocytes were positively selected by magnetically-activated cell sorting using mouse CD45 MicroBeads (Miltenyi Biotec). Purification of magnetically-activated sorted cells was confirmed by flow cytometry and was >95%.

Gating strategy

Mouse ILCs were defined as live, CD45+, lineage- (CD3, CD5, NK1.1, CD11b, CD11c, CD19, FcεR1), CD25+, CD127+ cells (Extended Data Fig. 1d, as previously described Ref. 7), mouse ILC2s were defined as live, CD45+, lineage-, CD25+, ST2+ cells (Extended Data Fig. 1d, as previously described Ref. 7). Central memory T cells were defined as live CD45+, CD3+, NK1.1-, CD8+, CD62l+, CD44+ (Extended Data Fig. 5c). Dendritic cells were defined as live CD45+, CD3-, NK1.1-, Gr1-, F4/80-, CD19-, CD11c+, MHC-II+ (Extended Data Fig. 6g).

Human ILCs were defined as live CD45+, lineage- (CD3, CD5, CD56, CD11b, CD11c, CD16, CD19, TCR?/?), FcεR1), CD25+, CD127+ cells as previously described (Extended Data Fig. 1a, Ref. 7).

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Gene expression and cell identity controlled by anaphase-promoting complex

<https://doi.org/10.1038/s41586-020-2034-1>

Received: 9 November 2018

Accepted: 1 January 2020

Published online: 19 February 2020

 Check for updates

Eugene Oh^{1,2,8}, Kevin G. Mark^{1,2,8}, Annamaria Moccia^{1,2,7}, Edmond R. Watson³, J. Rajan Prabu³, Denny D. Cha^{1,2}, Martin Kampmann^{4,5,6}, Nathan Gamarra⁴, Coral Y. Zhou⁴ & Michael Rape^{1,2,8}✉

Metazoan development requires the robust proliferation of progenitor cells, the identities of which are established by tightly controlled transcriptional networks¹. As gene expression is globally inhibited during mitosis, the transcriptional programs that define cell identity must be restarted in each cell cycle^{2–5} but how this is accomplished is poorly understood. Here we identify a ubiquitin-dependent mechanism that integrates gene expression with cell division to preserve cell identity. We found that WDR5 and TBP, which bind active interphase promoters^{6,7}, recruit the anaphase-promoting complex (APC/C) to specific transcription start sites during mitosis. This allows APC/C to decorate histones with ubiquitin chains branched at Lys11 and Lys48 (K11/K48-branched ubiquitin chains) that recruit p97 (also known as VCP) and the proteasome, which ensures the rapid expression of pluripotency genes in the next cell cycle. Mitotic exit and the re-initiation of transcription are thus controlled by a single regulator (APC/C), which provides a robust mechanism for maintaining cell identity throughout cell division.

The self-renewal of stem cells endows organisms with the capacity to establish or regenerate their many tissues, but the misregulation of self-renewal contributes to tumorigenesis, tissue degeneration or ageing⁸. Although tightly controlled transcriptional networks establish the identity of self-renewing stem cells during interphase¹, changes in chromatin architecture and the activity of transcription factors restrict the synthesis of messenger RNA (mRNA) during mitosis⁹. Stem cells must therefore restart their gene-expression programs each time they enter a new cell cycle^{4,5}, which is facilitated by promoter elements that remain unwound during mitosis², hypersensitive to DNase I^{2,10}, and accessible to RNA polymerase II and transcription factors such as the TATA-box binding protein TBP^{3,11–13}. How dividing cells retain hallmarks of interphase transcription to preserve their identity is incompletely understood.

APC/C sustains stem cell identity

To understand how pluripotency is preserved through cell division, we fused green fluorescent protein (GFP) to the *OCT4* (also known as *POU5F1*) locus of human embryonic stem (ES) cells. Diploid *OCT4*-GFP human ES cells responded to differentiation cues with an efficiency similar to that of their untagged counterparts (Extended Data Fig. 1a, b). Using lentiviral infection with pooled short hairpin (sh)RNAs, we depleted about 900 enzymes and effectors of ubiquitylation, which control cell division and differentiation¹⁴; propagated *OCT4*-GFP human ES cells in pluripotency medium, or briefly induced differentiation by neural conversion; and then deep-sequenced populations with low versus high levels of *OCT4*-GFP (Fig. 1a). shRNAs that decreased *OCT4*-GFP abundance in self-renewing human ES cells

target pluripotency factors, whereas shRNAs that sustained *OCT4*-GFP expression upon neural conversion deplete proteins that are needed for robust differentiation.

We recovered the positive-control *OCT4*, as well as known stem-cell E3 ligases such as DDB1, TRIM28 and UBR5^{15–17}, as pluripotency factors (Fig. 1b, Extended Data Fig. 1c). Consistent with the need for human ES cells to preserve genomic and proteomic integrity, we identified proteins involved in DNA repair (DDB1, RNF168 and USP7) and quality-control pathways (BAG6, HUWE1, PSMA1, PSMA6, UBR5 and UBXN7). Many of the enzymes of the latter pathways bind or produce K11/K48-branched ubiquitin chains¹⁸, which we confirmed in human ES cells (Extended Data Fig. 1d). Physiological pairs of E3 ligases and deubiquitylases (such as HUWE1 and USP7) clustered according to their opposing activities. Importantly, the APC2 subunit of APC/C was required for pluripotency, whereas the counteracting deubiquitylase USP44¹⁹ supported differentiation (Fig. 1b, Extended Data Fig. 1c, e). Other subunits of APC/C and APC/C-specific E2 enzymes scored as pluripotency factors, with *P* values that were slightly below our stringent screen cut-off (Extended Data Fig. 1c).

We confirmed that the depletion of subunits of APC/C, of the mitotic coactivator of APC/C (CDC20) or of APC/C-specific E2 enzymes inhibited human ES cell pluripotency, as revealed by decreased levels of *OCT4* and *NANOG* (Fig. 1c, Extended Data Fig. 2a–c). Although less pronounced than its effects on protein levels, depletion of APC2 reduced the abundance of *OCT4* and *NANOG* mRNA (Extended Data Fig. 2d). Human ES cells arrested in S phase and unable to enter mitosis did not require APC/C for pluripotency (Extended Data Fig. 2e), indicating that APC/C acts during cell division. However, it was unlikely that APC/C inhibition interfered with pluripotency simply by stalling mitotic

¹Howard Hughes Medical Institute, University of California at Berkeley, Berkeley, CA, USA. ²Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA, USA.

³Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany. ⁴Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA, USA. ⁵Institute for Neurodegenerative Diseases, University of California at San Francisco, San Francisco, CA, USA. ⁶Chan Zuckerberg Biohub, San Francisco, CA, USA.

⁷Present address: Berkeley Lights, Emeryville, CA, USA. ⁸These authors contributed equally: Eugene Oh, Kevin G. Mark. ✉e-mail: mrape@berkeley.edu

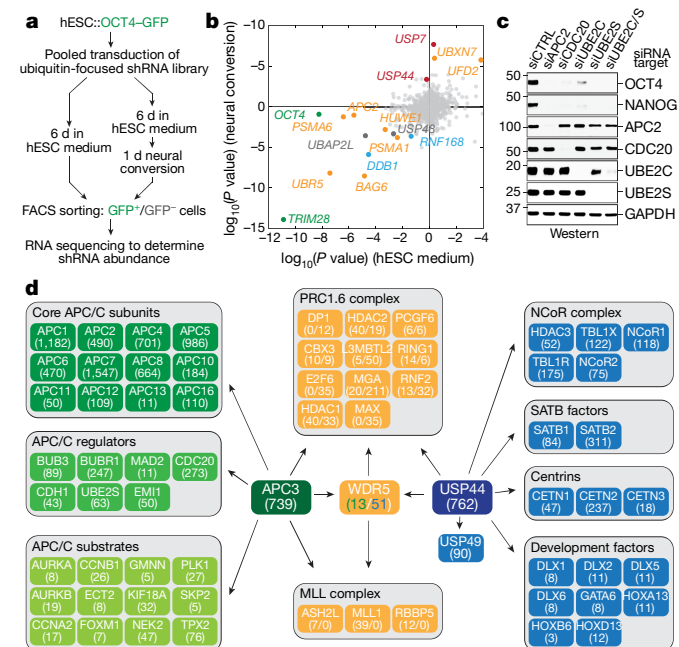


Fig. 1 | APC/C stabilizes human ES cell identity. a, Schematic of the ultracomplex shRNA screen. hESC, human ES cell. **b,** shRNA screen identifies genes that are important for pluripotency. Each dot ($n = 886$ unique genes) represents the P value of a gene (two-sided Mann–Whitney U test, not corrected for multiple hypothesis testing), calculated from comparing the collection of shRNAs that target each gene to all negative-control shRNAs measured in each subpopulation (low versus high levels of OCT4–GFP). Orange, genes that encode enzymes or effectors of K11/K48 branched-chain synthesis; red, genes that encode deubiquitylases that oppose K11/K48-specific E3 ligases; blue, genes that encode DNA-repair enzymes; and green, positive controls. *UFD2* is also known as *UBE4B*. Knockdown of genes indicated below and to the left of zero results in lower levels of OCT4–GFP; depletion of genes indicated above and to the right of zero maintains or increases the level of OCT4–GFP. **c,** Western blot of pluripotency markers upon APC/C-subunit knockdown in asynchronous H1 human ES cells. This experiment was performed five independent times with similar results. siCTRL, control small interfering (si) RNA; siAPC2, siCDC20, siUBE2C and siUBE2S denote siRNAs against *APC2*, *CDC20*, *UBE2C* and *UBE2S*, respectively. siUBE2C/S, siRNA against *UBE2C* and *UBE2S*. **d,** Interaction network of APC/C, WDR5 and USP44. Values listed in parentheses are total spectral counts of tryptic peptides of indicated proteins; values separated by a solidus denote proteins that coprecipitate with APC3 (left) or USP44 (right).

progression, as loss of the APC/C-specific E2 enzyme UBE2C diminished OCT4 and NANOG levels without affecting the G2/M population (Fig. 1c, Extended Data Fig. 2f). Collectively, these findings indicated that the essential mitotic regulator APC/C also helps to preserve the stem-cell state, identifying APC/C as a strong candidate for maintaining cell identity through cell division.

APC/C works with WDR5 in human ES cells

We speculated that the identification of APC/C or USP44 substrate adaptors required for pluripotency might point to ubiquitylated proteins that preserve human ES cell identity. Using mass spectrometry, we found that USP44—in addition to known partners—also engaged WDR5, a chromatin-associated factor that binds methylated histone H3K4 at active interphase promoters^{6,7,20} (Fig. 1d). Endogenous APC/C also interacted with WDR5 during mitosis (Fig. 1d), which we confirmed by reciprocal purification of WDR5 (Extended Data Fig. 3a). In addition, mitotic WDR5 bound the transcription factor TFIID (which includes TBP), as well as chromatin remodellers INO80 and CHD1 (Extended Data Fig. 3a).

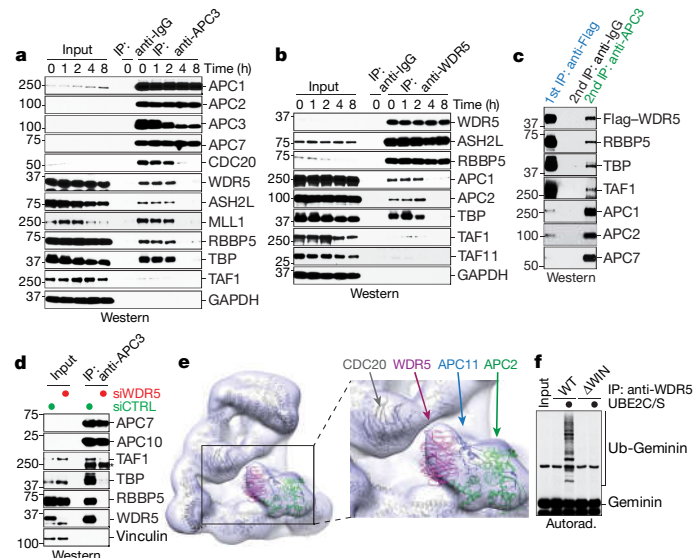


Fig. 2 | WDR5 is an APC/C substrate coadaptor. a, Immunoprecipitation (IP) of endogenous APC3 from HeLa cells reveals that APC/C binds WDR5 and TBP in mitosis. Prometaphase HeLa cells were released into fresh medium to restart the cell cycle. This experiment was performed three independent times with similar results. **b,** Immunoprecipitation of endogenous WDR5 from HeLa cells confirms that WDR5 associates with APC/C subunits and TBP in mitosis. This experiment was performed three independent times with similar results. **c,** Sequential immunoprecipitations of APC/C in complex with Flag-tagged WDR5 from mitotic HEK293T cells reveal that APC/C–WDR5 and TBP form a ternary complex. Flag–WDR5 was first purified from prometaphase cells, and next purified with anti-APC3. This experiment was performed once. **d,** Endogenous APC3 immunoprecipitations from control versus WDR5-depleted human ES cells show that the association of APC/C with TBP is bridged through WDR5. This experiment was performed twice with similar results. siWDR5, siRNA against *WDR5*. Asterisk denotes nonspecific band. **e,** The approximately 20 Å resolution negative-stain electron-microscopy model corroborates the association of WDR5 with the catalytic core of APC/C. **f,** Flag–WDR5 purified from mitotic HeLa cells contains active APC/C. Flag-tagged wild-type (WT) WDR5 or Flag–WDR5(Δ WIN) were purified from mitotic HeLa cells, and incubated with E1, UBE2C, UBE2S, ubiquitin, ATP and 35 S-labelled geminin. This experiment was performed two independent times with similar results. Autorad., autoradiography.

As with APC/C and TFIID–TBP²¹, depleting WDR5 diminished OCT4 and NANOG levels in human ES cells (Extended Data Fig. 3b). Human ES cells that are unable to enter mitosis did not require WDR5 for pluripotency (Extended Data Fig. 2e), which suggests that WDR5 acts during cell division. Consistently, the loss of WDR5 in human ES cells decreased the levels of K11-linked, as well as K11/K48-branched, ubiquitin chains—the mitotic products of APC/C¹⁸—to an extent similar to that seen after depletion of APC2 (Extended Data Fig. 3b). As in mouse ES cells²⁰, loss of WDR5 did not affect mitotic duration (Extended Data Fig. 3c), but codepletion of WDR5 and APC2 caused human ES cells to die shortly after exiting mitosis (Extended Data Fig. 3d–g). These findings suggested that WDR5 cooperates with APC/C to ensure human ES cell identity and survival, and does not impinge on the role of APC/C in controlling cell division.

Reciprocal immunoprecipitations of endogenous proteins from somatic cells showed that APC/C, WDR5, and TBP engage each other during early mitosis, when APC/C binds CDC20 (Fig. 2a, b). A similar mitotic increase in the interaction between APC/C and WDR5 was seen in human ES cells (Extended Data Fig. 3h). Sequential affinity purifications revealed that APC/C, WDR5 and TBP were part of the same complex (Fig. 2c), the formation of which depended on WDR5 (Fig. 2d). In contrast to APC/C, WDR5 engaged USP44 also during interphase (Extended Data Fig. 3i).

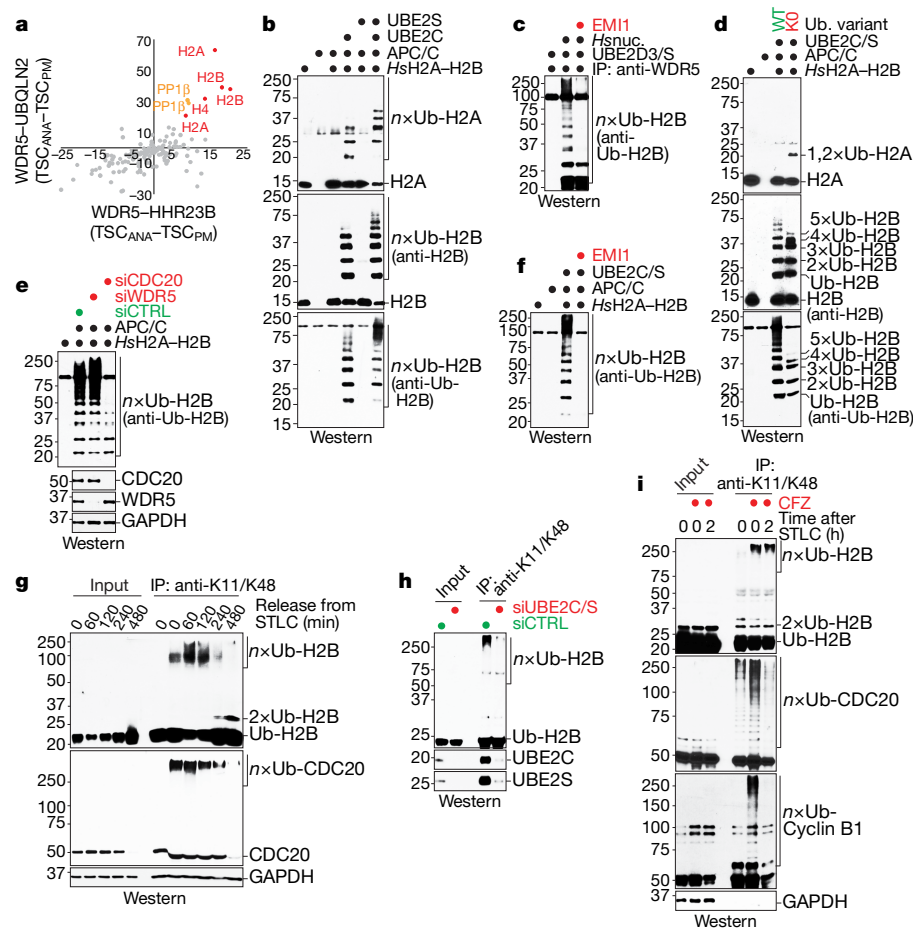


Fig. 3 | APC/C–WDR5 decorates histone proteins with K11/K48-branched ubiquitin chains. **a**, Mass spectrometry of WDR5–HHR23B and WDR5–UBQLN2 traps identifies histones as candidate substrates. Traps were affinity-purified from prometaphase (PM) or anaphase (ANA) HeLa cells with low or high APC/C activity, respectively. TSC, total spectral counts. **b**, APC/C–CDC20 purified from mitotic HeLa S3 cells ubiquitylates recombinant human (*Homo sapiens*, Hs) H2A–H2B dimers. This experiment was performed four independent times with similar results. **c**, APC/C–WDR5 ubiquitylates H2B in polynucleosomes (nuc.) purified from HeLa cells and is inhibited by the APC/C inhibitor EMI1. This experiment was performed three independent times with similar results. **d**, APC/C–WDR5 ubiquitylates multiple Lys residues in histones, as seen with Lys-free ubiquitin (K0). This experiment was performed two independent times with similar results. **e**, APC/C-dependent ubiquitylation of

H2B requires CDC20 in vitro. APC/C was purified from mitotic HeLa cells depleted of CDC20 or WDR5. This experiment was performed once. **f**, Ubiquitylation of H2B by APC/C is dependent on UBE2C and UBE2S, and inhibited by EMI1. This experiment was performed two independent times with similar results. **g**, Endogenous H2B is modified with K11/K48-branched chains, as seen by denaturing purification from synchronized HeLa cells. This experiment was performed three independent times with similar results. STLC, S-trityl-L-cysteine. **h**, Mitotic K11/K48 modification of endogenous H2B in human ES cells is dependent on UBE2C and UBE2S. This experiment was performed two independent times with similar results. **i**, Proteasome inhibition stabilizes mitotic K11/K48-modified H2B in H1 human ES cells. This experiment was performed two independent times with similar results. CFZ, carfilzomib.

WDR5 uses distinct surfaces to recognize WDR5-binding motifs (WBMs) and WDR5-interacting (WIN) motifs⁶. Disrupting the ability of WDR5 to bind WIN motifs (WDR5(ΔWIN)) blocked the association of WDR5 with APC/C and USP44, but not with TBP (Extended Data Figs. 3i, 4a–c). Accordingly, the compound MM-102—which targets the site on WDR5 that binds the WIN motif²²—prevented WDR5 from binding APC/C (Extended Data Fig. 4d), and WDR5(ΔWIN) did not sustain human ES cell pluripotency (Extended Data Fig. 4e). The ability of WDR5 to detect WBMs is not required for APC/C recognition, but is needed to bind TBP (Extended Data Fig. 4a, c).

Crosslinking experiments revealed that WDR5, but not WDR5(ΔWIN), binds APC/C close to CDC20 and the catalytic site that is composed of APC2 and APC11 (Extended Data Fig. 5a). Using in vitro translation, we identified APC2 as a specific binding partner of WDR5 (Extended Data Fig. 5b, c). We confirmed these findings by negative-stain electron microscopy, which showed that WDR5 is situated near CDC20 and docks against APC2 and APC11 (Fig. 2e).

Despite the proximity of WDR5 to the active site of APC/C, we could not detect APC/C-dependent ubiquitylation of WDR5 nor did excess WDR5 prevent the modification of APC/C substrates (Extended Data Fig. 6a, b). Instead, mitotic WDR5 complexes—which contain APC/C (Fig. 2b, Extended Data Fig. 3a)—supported the in vitro ubiquitylation of canonical APC/C substrates (Fig. 2f, Extended Data Fig. 6c). Mitotic WDR5 also coprecipitated K11-linked chains produced in cells (Extended Data Fig. 6d), which was dependent upon UBE2S (Extended Data Fig. 6e). We conclude that WDR5 binds active APC/C without being ubiquitylated itself, which suggests that WDR5 is a coadaptor that delivers APC/C to specific (probably chromatin-bound) substrates.

APC/C–WDR5 polyubiquitylates histones

To identify substrates of the APC/C–WDR5 complex, we used an approach that was previously established for SCF E3 ligases²³. We fused WDR5 to the ubiquitin-binding domains of HHR23B or UBQLN2, which

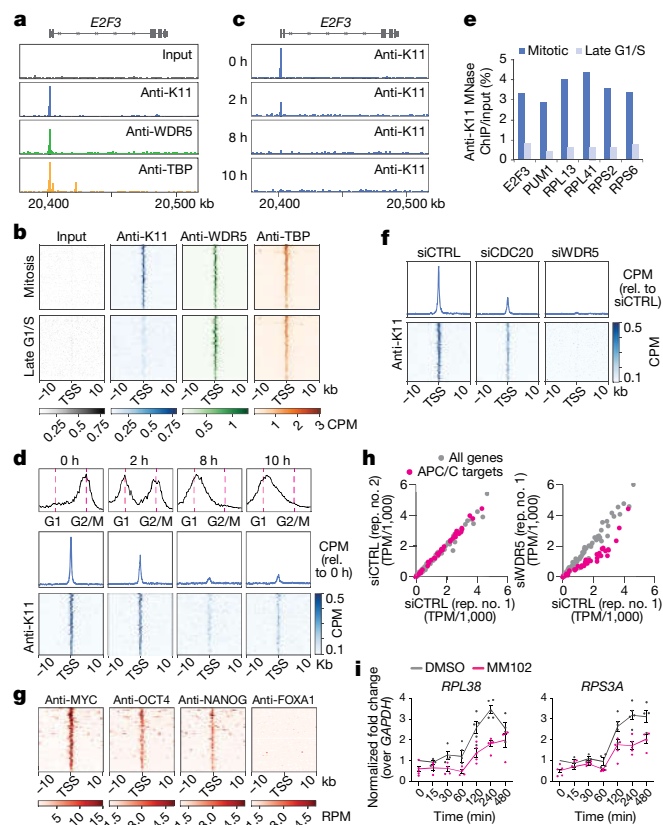


Fig. 4 | APC/C-dependent ubiquitylation occurs at TSSs of human ES cell genes. **a**, Genome browser track of *E2F3*. MNase ChIP-seq of indicated antibodies were performed from mitotic H1 human ES cells. **b**, K11 is deposited at select TSSs co-occupied by WDR5 in human ES cells. Heat map of co-occupied genes at TSSs from MNase ChIP-seq experiments of indicated antibodies. H1 human ES cells were collected after STLC treatment (mitosis) and after an 8-h release (late G1/S phase). CPM, counts per million. **c**, Genome browser track of *E2F3* from MNase ChIP-seq of anti-K11 in human ES cells throughout a mitotic release. **d**, Flow cytometry analysis of H1 human ES cells upon mitotic synchronization and release into fresh medium (top). Metagenome analysis of K11- and WDR5-occupied TSSs (middle). Heat map of individual K11- and WDR5-occupied TSSs from anti-K11 MNase ChIP-seq experiments throughout a mitotic release (bottom). **e**, Anti-K11 MNase ChIP-qPCR validates MNase ChIP-seq findings that K11 is deposited only during mitosis in H1 human ES cells. The same extract used in **c** was used for this experiment. **f**, Depletion of CDC20 or WDR5 causes robust depletion of K11 chains at select TSSs. **g**, MNase ChIP-seq from HUES64 human ES cells reveals that endogenous targets of APC/C-WDR5 are strongly enriched in binding sites for MYC, OCT4 and NANOG. **h**, Loss of APC/C-WDR5 function interferes with expression of genes marked with K11-linked chains in H1 human ES cells. Poly(A)⁺-selected RNA was purified from asynchronous H1 human ES cells transfected with control siRNA or siRNA against *WDR5* for 48 h and subjected to RNA sequencing. TPM, transcripts per million. **i**, Real-time qPCR analysis of nascent RNA reveals APC/C-WDR5 target genes are reactivated upon mitotic exit dependent on WDR5. Mitotic H1 human ES cells were treated with or without 50 μ M MM102 and supplemented with 20 μ M Z-VAD-FMK. Cells were released into fresh medium containing DMSO or 50 μ M MM102. Real-time qPCR experiments were performed with oligonucleotides spanning intron-exon junctions. Values represent the mean of independent replicates \pm s.e.m. ($n = 3$ for $t = 15$ min, $n = 4$ for $t = 30, 60$ and 480 min and $n = 5$ for $t = 0, 120$ and 240 min).

detect K11/K48-branched chains produced by APC/C¹⁸, and purified both constructs under conditions of low or high APC/C activity. Ubiquitylated substrates were expected to be trapped by both fusions in cells with active APC/C. These experiments identified histones as likely APC/C-WDR5 substrates (Fig. 3a).

In vitro reconstitution using human histone H2A-H2B dimers and H3-H4 tetramers, or *Xenopus laevis* H2A-H2B dimers and octamers, revealed efficient APC/C-dependent ubiquitylation of H2A, H2B and H3, but not of H4 (Fig. 3b, Extended Data Fig. 7a-c). H2A-H2B dimers, octamers and polynucleosomes were also strongly ubiquitylated by WDR5-bound APC/C and by endogenous APC/C purified from human ES cells (Fig. 3c, Extended Data Fig. 7b-d). Histone polyubiquitylation occurred at multiple sites (Fig. 3d), including K120 of H2B—the monoubiquitylation of which leads to transcriptional activation, and is negatively regulated by USP44²⁴.

In contrast to mitotic APC/C, APC/C obtained from asynchronous or S-phase cells did not modify histones (Extended Data Fig. 7e). APC/C-dependent polyubiquitylation of histones was also blocked by the depletion of CDC20 (the mitotic coactivator of APC/C), by the addition of the APC/C inhibitor EMI1 or mutation of the K11 of ubiquitin (Fig. 3e, f, Extended Data Fig. 7f, g). H2B ubiquitylation was outcompeted by a canonical APC/C substrate, but less so by a D-box mutant substrate (Extended Data Fig. 7h), which indicates that histones are recognized by the D-box coreceptor composed of CDC20 and APC10²⁵.

Denaturing purifications of K11/K48-branched chains revealed abundant ubiquitylation of endogenous H2B during early mitosis, at a time when CDC20 is decorated with such conjugates (Fig. 3g). Underscoring the role of APC/C, H2B modification with K11/K48-linked chains was strongly reduced by UBE2C and UBE2S depletion (Fig. 3h). Ubiquitylated H2B accumulated upon proteasome inhibition (Fig. 3i, Extended Data Fig. 7i), consistent with K11/K48-branched conjugates targeting proteins for degradation^{18,26}. We conclude that APC/C-WDR5 modifies multiple histones with K11/K48-branched ubiquitin chains during mitosis.

APC/C acts at transcription start sites

As total histone levels did not drop during mitotic exit (Extended Data Fig. 8a), we hypothesized that APC/C-WDR5 targets histones at select chromosome locations. To identify this population, we performed genome-wide micrococcal-nuclease chromatin immunoprecipitation with sequencing (MNase ChIP-seq) analysis of K11-linked chains, WDR5 and TBP in prometaphase human ES cells. Because the vast majority of K11 linkages are assembled during mitosis by APC/C^{18,27}, tracking this type of chain enabled us to monitor APC/C even if it interacted with its targets only transiently. MNase was used, as sonication fragmented polymeric ubiquitin chains and reduced the specific ChIP-seq signal (Extended Data Fig. 8b).

Notably, K11-linked and K11/K48-branched chains (that is, active APC/C) accumulated at specific genes in mitotic human ES cells that were co-occupied by WDR5 and TBP (Fig. 4a, b, Extended Data Fig. 8c-e). Chromatin-bound K11-linked chains were abundant during early mitosis (when APC/C is activated by CDC20), but were undetectable during late G1 or early S phase, when APC/C is inactive (Fig. 4c-e). By contrast, WDR5 and TBP were found at these promoters throughout the cell cycle (Fig. 4b). Depletion of CDC20, UBE2S or WDR5, and chemical inhibition of WDR5, strongly reduced K11-linked chains at APC/C-WDR5 target genes (Fig. 4f, Extended Data Fig. 8f, g). By heterologous expression of CDC20 and WDR5, we showed that mitotic APC/C-WDR5 also associated with specific genes in somatic cells (Extended Data Fig. 8h).

The majority of APC/C-WDR5 target sites were within 100 base pairs of the transcription start site (TSS); this location contains TBP-binding sites, as we confirmed for select targets by ChIP with quantitative PCR (ChIP-qPCR) (Extended Data Fig. 8i, j). Gene ontology (GO) analyses revealed that most APC/C-WDR5 target genes encode proteins that are involved in ribosome function (GO: 0003735, $P = 1.2 \times 10^{-56}$) and mRNA translation (GO: 0006413, $P = 2.2 \times 10^{-59}$). These genes are among the very first to be expressed upon mitotic exit⁴, dependent upon WDR5 and MYC^{6,28}. Accordingly, APC/C-WDR5 target genes were strongly bound by the stem-cell transcription factors MYC, OCT4 and NANOG

(Fig. 4g, Extended Data Fig. 8k), whereas transcription factors linked to differentiation did not accumulate at these sites (Extended Data Fig. 9). When we compared the set of APC/C–WDR5 target genes from HEK293T cells with gene-expression profiles, we noticed strong overlaps with human ES cell lines (Extended Data Fig. 10a).

Given the enrichment of APC/C–WDR5 at the TSSs of pluripotency genes and the requirement for this complex for self-renewal, we asked whether APC/C–WDR5 controls the transcription of its target genes. Notably, depletion of WDR5 strongly downregulated only those genes that were marked by K11-linked chains, WDR5 and TBP during mitosis (Fig. 4h, Extended Data Fig. 10b, c). Real-time qPCR analyses of nascent mRNAs using oligonucleotides that span intron–exon junctions showed that APC/C–WDR5 target genes were expressed immediately upon mitotic exit, dependent on WDR5 (Fig. 4i, Extended Data Fig. 10d). APC/C–WDR5 target genes are expressed at high levels (Extended Data Fig. 10e), and hence, particularly reliant on rapid reactivation after mitosis. Polyubiquitylation by APC/C–WDR5 therefore promotes early post-mitotic expression of genes controlled by stem cell transcription factors.

APC/C recruits p97 and the proteasome

Consistent with K11/K48-branched chains recruiting the cellular degradation machinery^{18,26}, the p97 adaptor UBXN7 and proteasome subunits scored in our screen (Fig. 1b). The p97–UBXN7 complex captured K11/K48-modified H2B *in vitro* (Extended Data Fig. 10f) and strongly bound K11/K48-ubiquitylated H2B in cells (Extended Data Fig. 10g). Moreover, p97 inhibition by NMS-873 caused the same strong increase in K11/K48-ubiquitylation of H2B as seen with proteasome inhibition (Extended Data Fig. 10h). Both MNase ChIP–seq and ChIP–qPCR experiments revealed that p97 and the proteasome were required for the loss of ubiquitylated proteins from the TSSs of APC/C–WDR5 target genes upon mitotic exit (Extended Data Fig. 10i, j). These findings suggest that APC/C–WDR5 might act by destabilizing histones at specific TSSs during mitosis.

Discussion

Our findings reveal a mechanism for how cell identity is preserved through cell division (Extended Data Fig. 10k). WDR5 and TBP bind promoters of genes transcribed in interphase. When cells enter mitosis, WDR5 and TBP remain associated with their targets but, instead of recruiting RNA polymerase II, they deliver APC/C to TSSs demarcated by the pluripotency factors MYC, OCT4 and NANOG. At these TSSs, APC/C decorates histones with K11/K48-branched chains, which attract p97 and the proteasome. We propose that subsequent histone degradation opens the TSSs for the rapid postmitotic expression of pluripotency genes. As it also triggers mitotic exit²⁹, APC/C therefore tightly coordinates cell division and gene-expression regulation.

The newly identified cofactor WDR5 binds APC/C through the same surface as it uses to engage the MLL1 methyltransferase, another regulator of postmitotic gene expression³⁰. Histone methylation might strengthen the interaction of WDR5 with promoters, which could facilitate subsequent recruitment of APC/C. WDR5 also engages OCT4, MYC and TFIID–TBP, all of which bind APC/C–WDR5 target genes and have vital roles in mitotic bookmarking. WDR5 thus appears to orchestrate distinct steps in the regulation of mitotic gene expression by mediating the recruitment of transcription factors, histone methylation and nucleosome destabilization.

Partial APC/C inhibition in neural progenitors triggered cell differentiation similar to that noted upon loss of APC/C–WDR5 in human ES cells³¹. Conversely, cellular reprogramming and somatic-cell nuclear transfer are more efficient during mitosis^{32,33}, at times that coincide with APC/C–WDR5-dependent histone ubiquitylation. This further implies a role for APC/C–WDR5 in pluripotency control, which comes with practical implications: if APC/C–WDR5 acts in cancer stem cells as in human ES cells, combinations of APC/C and WDR5 inhibitors might

impede the self-renewal of disease-driving cell populations and should be tested for their efficiency in cancer therapy.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2034-1>.

- Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
- Michelotti, E. F., Sanford, S. & Levens, D. Marking of active genes on mitotic chromosomes. *Nature* **388**, 895–899 (1997).
- Teves, S. S. et al. A stable mode of bookmarking by TBP recruits RNA polymerase II to mitotic chromosomes. *eLife* **7**, e35621 (2018).
- Palozola, K. C. et al. Mitotic transcription and waves of gene reactivation during mitotic exit. *Science* **358**, 119–122 (2017).
- Hsiung, C. C. et al. A hyperactive transcriptional state marks genome reactivation at the mitosis–G1 transition. *Genes Dev.* **30**, 1423–1439 (2016).
- Thomas, L. R. et al. Interaction with WDR5 promotes target gene recognition and tumorigenesis by MYC. *Mol. Cell* **58**, 440–452 (2015).
- Wysocka, J. et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Keyes, B. E. & Fuchs, E. Stem cells: aging and transcriptional fingerprints. *J. Cell Biol.* **217**, 79–92 (2018).
- Prescott, D. M. & Bender, M. A. Synthesis of RNA and protein during mitosis in mammalian tissue culture cells. *Exp. Cell Res.* **26**, 260–268 (1962).
- Martinez-Balbás, M. A., Dey, A., Rabindran, S. K., Ozato, K. & Wu, C. Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell* **83**, 29–38 (1995).
- Caravaca, J. M. et al. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev.* **27**, 251–260 (2013).
- Festuccia, N. et al. Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol.* **18**, 1139–1148 (2016).
- Kadaue, S. et al. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* **150**, 725–737 (2012).
- Rape, M. Ubiquitylation at the crossroads of development and disease. *Nat. Rev. Mol. Cell Biol.* **19**, 59–70 (2018).
- Buckley, S. M. et al. Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell* **11**, 783–798 (2012).
- Gao, J. et al. The CUL4–DDB1 ubiquitin ligase complex controls adult and embryonic stem cell differentiation and homeostasis. *eLife* **4**, e07539 (2015).
- Hu, G. et al. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23**, 837–848 (2009).
- Yau, R. G. et al. Assembly and function of heterotypic ubiquitin chains in cell-cycle and protein quality control. *Cell* **171**, 918–933 (2017).
- Stegmeier, F. et al. Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities. *Nature* **446**, 876–881 (2007).
- Ang, Y. S. et al. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–197 (2011).
- Pijnappel, W. W. et al. A central role for TFIID in the pluripotent transcription circuitry. *Nature* **495**, 516–519 (2013).
- Karatas, H. et al. High-affinity, small-molecule peptidomimetic inhibitors of MLL1/WDR5 protein–protein interaction. *J. Am. Chem. Soc.* **135**, 669–682 (2013).
- Mark, K. G., Loveless, T. B. & Toczyński, D. P. Isolation of ubiquitinated substrates by tandem affinity purification of E3 ligase-polyubiquitin-binding domain fusions (ligase traps). *Nat. Protoc.* **11**, 291–301 (2016).
- Fuchs, G. et al. RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. *Mol. Cell* **46**, 662–673 (2012).
- Chang, L. F., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Molecular architecture and mechanism of the anaphase-promoting complex. *Nature* **513**, 388–393 (2014).
- Meyer, H. J. & Rape, M. Enhanced protein degradation by branched ubiquitin chains. *Cell* **157**, 910–921 (2014).
- Matsumoto, M. L. et al. K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. *Mol. Cell* **39**, 477–484 (2010).
- Aho, E. R. et al. Displacement of WDR5 from chromatin by a WIN site inhibitor with picomolar affinity. *Cell Rep.* **26**, 2916–2928 (2019).
- King, R. W. et al. A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. *Cell* **81**, 279–288 (1995).
- Blöbel, G. A. et al. A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Mol. Cell* **36**, 970–983 (2009).
- Pilaz, L. J. et al. Prolonged mitosis of neural progenitors alters cell fate in the developing brain. *Neuron* **89**, 83–99 (2016).
- Halley-Stott, R. P., Jullien, J., Pasque, V. & Gurdon, J. Mitosis gives a brief window of opportunity for a change in gene transcription. *PLoS Biol.* **12**, e1001914 (2014).
- Egli, D., Birkhoff, G. & Eggan, K. Mediators of reprogramming: transcription factors and transitions through mitosis. *Nat. Rev. Mol. Cell Biol.* **9**, 505–516 (2008).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Mammalian cell culture

Human embryonic kidney (HEK)293T and HeLa cells were maintained in DMEM plus 10% fetal bovine serum. Plasmid transfections were performed using polyethylenimine (PEI) at a 1:3 ratio of DNA (in μg) to PEI (in μl at a 1 mg ml^{-1} stock concentration). siRNA transfections were performed using 40 nM of indicated siRNAs and a 1:400 dilution of RNAiMAX transfection reagent (Thermo Fisher, 13778150). Lentiviruses were produced in HEK293T cells by cotransfection of lentiviral and packaging plasmids using Lipofectamine 2000 transfection reagent (Thermo, 11668027). Viruses were collected 48 h after transfection, concentrated using the Lenti-X concentrator (Takara, 631232), aliquoted, and stored at -80°C for later use. HEK293T cells were purchased directly from the Berkeley Cell Culture Facility (authenticated by short tandem repeat analysis). HeLa cells were not authenticated.

Human ES cells (WiCell, WA01/H1) were grown in mTeSR1 medium (StemCell Technologies, 85850) on human-ES-cell-qualified Matrigel-coated plates (Corning, 354277) with daily medium change. H1 cells were passaged by collagenase (StemCell Technologies, 07909) for routine maintenance or accutase (StemCell Technologies, 07920) for siRNA transfections, lentiviral infections or when single cells were required. For siRNA transfections, single-cell suspensions of H1 cells were generated by accutase treatment and $2\text{--}5 \times 10^5$ cells were seeded on a Matrigel-coated well of a 6-well plate with 1.8 ml of mTeSR1 containing $10\text{ }\mu\text{M}$ of Y-27632 (StemCell Technologies, 72308) and a 0.2 ml mixture of indicated siRNAs (at a final concentration of 40 nM) and a 1:400 dilution of RNAiMAX transfection reagent buffered in Opti-MEM. For lentiviral infections, single-cell suspensions of H1 cells were generated by accutase treatment and $1.5\text{--}3 \times 10^5$ cells were seeded on a Matrigel-coated well of a 6-well plate with 2 ml of mTeSR1 containing $10\text{ }\mu\text{M}$ of Y-27632, polybrene (at a final concentration of $6\text{ }\mu\text{g ml}^{-1}$), and lentiviruses produced from HEK293T cells for 2 h. The medium was immediately exchanged with 2 ml of fresh mTeSR1 containing $10\text{ }\mu\text{M}$ of Y-27632 only. Human ES cells were drug-selected 24–48 h after infection. H1 cells were positive for OCT4 and NANOG expression and karyotype analysis showed no chromosomal anomalies.

All cell lines were routinely tested for mycoplasma contamination using the MycoAlert mycoplasma detection kit (Lonza, LT07-218). All cell lines tested negative for mycoplasma.

Generation of OCT4-eGFP-P2A-PURO^R human ES cells

The OCT4 locus was targeted for gene editing in H1 cells by TALE nucleases as previously described³⁴. An in-frame fusion, consisting of enhanced GFP (eGFP) followed by the self-cleaving P2A peptide and the puromycin resistance gene (puromycin *N*-acetyltransferase), was generated at the C terminus of the OCT4 locus. In brief, single-cell suspensions of H1 cells were generated by accutase treatment and 1×10^7 cells were resuspended in ice-cold $1 \times \text{PBS}$ with $40\text{ }\mu\text{g}$ of the DONOR plasmid and $5\text{ }\mu\text{g}$ each of the TALEN plasmids (T4 and T8). Cells were electroporated in a 0.4-cm cuvette at 250 V and 500 μF with the Gene Pulser II electroporating system (Bio-Rad). Electroporated cells were immediately resuspended in mTeSR1, washed to remove lysed debris and seeded on 2 Matrigel-coated 15-cm plates in mTeSR1 containing $10\text{ }\mu\text{M}$ of Y-27632. H1 cells were selected for 10–14 days with puromycin (at a final concentration of $0.5\text{ }\mu\text{g mg}^{-1}$) 72 h after electroporation. Colonies were manually scored and transferred to fresh plates. A single allele of the OCT4 locus was fused with the eGFP-P2A-PURO^R cassette as verified by Southern blot analysis (data not shown). Karyotype analysis was performed by WiCell.

Neural conversion of human ES cells

Neural induction of human ES cells were performed as previously described³⁵, using STEMdiff Neural Induction Medium (StemCell Technologies, 05839). Single-cell suspensions of H1 cells were generated by accutase treatment and 1.5×10^6 cells were seeded in a well of 6-well plate with 4 ml of STEMdiff neural induction medium containing $10\text{ }\mu\text{M}$ Y-27632. Cells were treated with daily medium changes, and collected when indicated.

Ultracomplex shRNA screen

The shRNA library was constructed as previously described³⁶. In brief, the shRNA library was divided into four sublibraries, cloned into lentiviral expression vectors and transfected into HEK293T cells with TransIT-293 transfection reagent (Mirus, MIR 2700) for virus production. Human ES cells were infected with lentiviruses overnight and cultured in mTeSR1 for six days or in mTeSR1 for six days followed by STEMdiff neural induction medium for one day. Human ES cells were then sorted by fluorescence-activated cell sorting using an INFLUX cell sorter (BD) at the Flow Cytometry Core Facility at UC Berkeley. Cells were sorted on the basis of the strength of their GFP expression into three populations. Sequencing libraries were prepared from sorted cells as previously described³⁶, sequenced on a HiSeq 2000 (Illumina) and analysed using previously described scripts³⁶.

Cell synchronization

HeLa cells were first synchronized in S phase by addition of thymidine (at a final concentration of 2 mM) for 24 h. S-phase cells were washed with $1 \times \text{PBS}$ to remove excess thymidine and released into fresh medium (DMEM/10%FBS) for 3 h. To arrest cells in prometaphase, released cells were treated with STLC (Sigma, 164739) (at a final concentration of $5\text{ }\mu\text{M}$) for 12–14 h. Finally, prometaphase cells were collected by vigorous pipetting, washed with $1 \times \text{PBS}$ and used for downstream applications, including immunoprecipitation assays and/or western blot analyses, or frozen in liquid nitrogen and stored at -80°C for later use. For cell-cycle studies, prometaphase cells were released into fresh medium and collected at the indicated time points. For drug inhibition studies, cells were released into medium containing $2\text{ }\mu\text{M}$ carfilzomib (Selleck, PR-171), $20\text{ }\mu\text{M}$ (R)-MG132 (Cayman, 13697) and/or $10\text{ }\mu\text{M}$ NMS-873 (Sigma, SML1128) for indicated times. For depletion studies, HeLa cells were transfected with 40 nM of indicated siRNAs and a 1:400 dilution of RNAiMAX transfection reagent (Thermo Fisher, 13778150) 24 h before synchronization.

Mitotic enrichment of HEK293T cells and H1 cells was achieved by adding STLC (at a final concentration of $5\text{ }\mu\text{M}$) to the culture medium for 14–16 h.

Purification of APC/C and APC/C–WDR5 complexes

Human APC/C and APC/C–WDR5 complexes were purified from HeLa extracts synchronized in prometaphase (see ‘Cell synchronization’). To purify APC/C–WDR5, HeLa cells were first PEI-transfected with $5\text{ }\mu\text{g}$ of pCMV 3×Flag–WDR5 (per 15-cm plate) for 24 h before synchronization. Collected prometaphase pellets were lysed in lysis buffer (20 mM HEPES, pH 7.4, 5 mM KCl, 150 mM NaCl, 1.5 mM MgCl_2 , 0.1% Nonidet P-40, $1 \times \text{cComplete}$ protease inhibitor cocktail (Roche, 04693159001) and $1\text{ }\mu\text{l}$ of benzonase (Millipore, 70746) per 15-cm plate). Detergent lysed cells were then subjected to a high-speed spin (20,000g) to remove cellular debris and the clarified extract was precleared with protein G-agarose resin (Roche, 11719416001). APC/C was purified with anti-CDC27 antibody (sc-9972, SCBT) precoupled to protein G-agarose resin for 3 h at 4°C , and APC/C–WDR5 was purified with anti-Flag M2 affinity resin (Sigma, A2220) for 1.5 h at 4°C . APC/C-coupled beads were washed $5 \times$ with lysis buffer (minus inhibitors and benzonase) before use.

Purification of recombinant proteins

WDR5 and WDR5(Δ WIN) were cloned into a pMAL expression vector containing a C-terminal 6 \times His tag and expressed in *Escherichia coli* BL21-CodonPlus (DE3)RIL cells. Transformed cells were grown at 37 °C to an optical density at 600 nm (OD₆₀₀) of 0.5 in LB broth containing 100 μ g ml⁻¹ ampicillin, 34 μ g ml⁻¹ chloramphenicol and 0.2% glucose, chilled on ice for 30 min, induced with 100 μ M isopropyl β -D-1-thiogalactopyranoside (IPTG) for 6 h at 16 °C, and collected by centrifugation. Collected cells were resuspended with lysis buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 2 mM 2-mercaptoethanol (BME), 1 mM EDTA, 10% glycerol, 0.2 mg ml⁻¹ lysozyme, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.1% Triton X-100), incubated on ice for 30 min, sonicated and clarified by high-speed centrifugation. The clarified extract was supplemented with 20 mM imidazole and bound to Ni-NTA resin (Qiagen, R90110) (2 ml of slurry per 1 l of bacterial culture) for 1 h at 4 °C. The resin was then washed 5 \times with wash buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 2 mM BME, 1 mM EDTA, 10% glycerol and 20 mM imidazole) and eluted 2 \times with elution buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 2 mM BME, 1 mM EDTA, 10% glycerol and 300 mM imidazole). The elutions were pooled, dialysed overnight in dialysis buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 2 mM BME, 1 mM EDTA and 10% glycerol), concentrated, aliquoted, snap-frozen in liquid nitrogen and stored at -80 °C for later use.

Securin and its variants were cloned into a pET28 expression vector containing an N-terminal 6 \times His tag followed by a TEV-protease cleavage site and expressed in LOBSTR BL21(DE3)-RIL cells. Transformed cells were grown at 37 °C to an OD₆₀₀ of 0.5 in LB broth containing 100 μ g ml⁻¹ ampicillin and 34 μ g ml⁻¹ chloramphenicol, chilled on ice for 30 min and induced with 100 μ M IPTG for 14–16 h at 16 °C. Induced cells were centrifuged, resuspended in lysis buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 2 mM BME, 10% glycerol, 0.2 mg ml⁻¹ lysozyme, 1 mM PMSF and 0.1% Triton X-100), incubated on ice for 30 min, sonicated and clarified by high-speed centrifugation. The clarified extract was supplemented with 20 mM imidazole and bound to Ni-NTA resin (2 ml of slurry per 1 l of culture) for 1 h at 4 °C. The resin was then washed 5 \times with wash buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 2 mM BME, 10% glycerol, 0.1% Triton X-100 and 20 mM imidazole) and eluted by TEV cleavage. The eluate was desalted using a PD10 column, concentrated, aliquoted, snap-frozen and stored at -80 °C for later use.

p97 was cloned into a pMAL expression vector and expressed in BL21-CodonPlus (DE3)RIL cells. Transformed cells were grown at 37 °C to an OD₆₀₀ of 0.5 in LB broth containing 100 μ g ml⁻¹ ampicillin and 34 μ g ml⁻¹ chloramphenicol, chilled on ice for 30 min and induced with 0.5 mM IPTG overnight at 18 °C. Induced cells were centrifuged, resuspended in lysis buffer (20 mM Tris 7.4, 300 mM NaCl, 5% glycerol, 0.2 mg ml⁻¹ lysozyme, 1 mM PMSF and 0.1% Triton X-100), incubated on ice for 30 min, sonicated and clarified by high-speed centrifugation. The clarified extract was bound to amylose resin (NEB, E8021) (2 ml of slurry per 1 l of culture) for 45 min at 4 °C. The resin was then washed 3 \times with 1 \times PBS, resuspended in 1 \times PBS containing 2 mM DTT, and stored at 4 °C for up to 1 month. Recombinant 6 \times His-p47 (also known as NSFL1C) and 6 \times His-UBXN7 were purified using previously described methods¹⁸.

In vitro transcription and translation

All in vitro synthesized substrates were cloned under the SP6 promoter. The corresponding plasmids can be found in Supplementary Table 1. ³⁵S-labelled substrates were generated by incubating 3 μ l (400 ng) of plasmid DNA in 20 μ l of rabbit reticulocyte lysate (Promega, L2080) supplemented with 2 μ l of ³⁵S-Met (PerkinElmer, NEG009H001MC) for 1 h at 30 °C. Reactions were terminated by rapid dilution with 1 \times PBS. ³⁵S-labelled substrates were used for in vitro ubiquitylation assays and/or MBP binding studies.

In vitro ubiquitylation

In vitro ubiquitylation assays were performed in a 10 μ l reaction volume: 0.25 μ l of 10 μ M E1 (250 nM final), 1 μ l of 10 μ M UBE2C (1 μ M final), 1 μ l of 10 μ M UBE2S (1 μ M final), 1 μ l of 10 mg ml⁻¹ ubiquitin (1 mg ml⁻¹ final) (Boston Biochem, U-100H), 1 μ l of 100 mM DTT, 1.5 μ l of energy mix (150 mM creatine phosphate, 20 mM ATP, 20 mM MgCl₂, 2 mM EGTA, pH to 7.5 with KOH), 2.25 μ l of 1 \times PBS, 1 μ l of 10 \times ubiquitylation assay buffer (250 mM Tris 7.5, 500 mM NaCl, and 100 mM MgCl₂) and 3 μ l of substrate (in vitro translated or recombinant) were premixed and added to 5 μ l of APC/C- or APC/C-WDR5-purified bed resin (see 'Purification of APC/C and APC/C-WDR5 complexes'). Reactions were performed at 30 °C with shaking for 30 min, unless noted otherwise. Reactions were stopped by adding 2 \times urea sample buffer and resolved on SDS-acrylamide gels. E1, UBE2C and UBE2S were purified as previously described²⁶. Recombinant human H2A-H2B dimers (NEB, M2508S), recombinant *X. laevis* H2A-H2B dimers and octamers, recombinant human H3-H4 tetramers (NEB, M2509S), or purified human nucleosomes (EpiCypher, 16-0003) were used at a final concentration of 500 nM.

MBP binding studies

For in vitro transcription and translation binding assays, 10 μ l of ³⁵S-labelled substrate was diluted down to 400 μ l with prechilled 1 \times PBS containing 0.1% Nonidet P-40 and mixed with 2 μ l of 1 mg ml⁻¹ of MBP-fused bait (see 'Purification of recombinant proteins') and 8 μ l of amylose slurry (NEB, E8021). The binding was performed for 2 h at 4 °C with mixing, and the amylose resin was subsequently washed 3 \times with 1 \times PBS. The bound prey was eluted with 2 \times urea sample buffer, resolved on an SDS-acrylamide gel and visualized by a Typhoon scanner.

For coadaptor-bound p97 binding studies, coadaptor-bound p97 resin was made by mixing 0.1 ml of p97-coupled amylose slurry (see 'Purification of recombinant proteins') with 0.2 ml of recombinant 6 \times His-p47 or 6 \times His-UBXN7 and 0.3 ml of 1 \times PBS containing 4 mM DTT for 45 min at 4 °C. The resin was washed 3 \times with 1 \times PBS containing 2 mM DTT and stored at 4 °C for up to 2 weeks. Ubiquitylated H2A-H2B dimers (see 'In vitro ubiquitylation') were added to 6 μ l of coadaptor-bound p97 slurry brought up in 0.6 ml of 1 \times PBS, incubated for 20 min at 4 °C, washed 5 \times with 1 \times PBS, eluted with 2 \times urea sample buffer and resolved on an SDS-acrylamide gel.

Crosslinking studies

APC/C complexes were first purified from HeLa cells synchronized in prometaphase. Before crosslinking, a 200 μ M working stock of the sulfhydryl-reactive and homobifunctional crosslinker 1,4-bis-maleimidobutane (BMB) was prepared in DMSO and a 20 μ M solution of recombinant MBP-WDR5 was pretreated with tris(2-carboxyethyl) phosphine (TCEP) (at a final concentration of 1 mM) in a 20 μ l reaction volume. Ten microlitres of purified APC/C slurry (see 'Purification of APC/C and APC/C-WDR5 complexes') was mixed with TCEP-treated MBP-WDR5 (at a final concentration of 2 μ M) and BMB (at a final concentration of 20 μ M) and incubated for 30 min at 22 °C with shaking. Reactions were stopped by adding 2 \times urea sample buffer and resolved on SDS-acrylamide gels.

K11/K48 denaturing immunoprecipitations

Denaturing K11/K48-linked ubiquitin immunoprecipitations were performed from cells arrested in prometaphase. Three 15-cm plates of confluent cells were collected and lysed in equal pellet volume with urea lysis buffer (20 mM Tris 7.5, 135 mM NaCl, 10% glycerol, 8 M urea, 1% Triton X-100, 5 μ M carfilzomib (Selleck, PR-171), 10 mM N-ethylmaleimide (NEM), 1 \times phosSTOP (Roche, 4906837001) and 1 \times cComplete protease inhibitor cocktail (Roche, 04693159001)), rotated for 1 h at room temperature, sonicated with a microtip sonicator (15 pulses at 50 amps), diluted 2-fold in dilution buffer (20 mM Tris 7.5, 135 mM NaCl, 10% glycerol, 5 μ M carfilzomib, 10 mM NEM, 1 \times phosSTOP and

1× cOmplete protease inhibitor cocktail) and clarified for 5 min at low speed (2,400g). Clarified extracts were incubated with 20 µg of anti-K11/K48 bispecific ubiquitin antibody or control normal mouse IgG and 40 µl of protein G-agarose slurry for 3 h at room temperature. Beads were washed 10× with dilution buffer, eluted with 2× urea sample buffer, and resolved on SDS-acrylamide gels.

Mass spectrometry

Mass spectrometry was performed on immunoprecipitates prepared from HEK293T cells. In brief, 20 15-cm plates of HEK293T cells were PEI-transfected (if indicated), grown to confluence, synchronized (if indicated), collected and lysed in lysis buffer (20 mM HEPES, pH 7.4, 5 mM KCl, 150 mM NaCl, 1.5 mM MgCl₂, 0.1% Nonidet P-40 and 1× cOmplete protease inhibitor cocktail). Lysed extracts were clarified by high-speed centrifugation, precleared with protein G-agarose slurry and bound to indicated antibodies pre-coupled to protein G-agarose resin (for immunoprecipitations of endogenous proteins) or anti-Flag M2 affinity resin (for immunoprecipitations of overexpressed proteins). Immunoprecipitates were then washed and eluted 3× at 30 °C with 0.5 mg ml⁻¹ of 3× Flag peptide (Sigma, F4799) buffered in 1× PBS plus 0.1% Triton X-100. Elutions were pooled and precipitated overnight at 4 °C with 20% trichloroacetic acid. Immunoprecipitates were then pelleted, washed 3× with an ice-cold acetone/0.1 N HCl solution, dried, resolubilized in 8 M urea buffered in 100 mM Tris 8.5, reduced with TCEP (at a final concentration of 5 mM) for 20 min, alkylated with iodoacetamide (at a final concentration of 10 mM) for 15 min, diluted 4-fold with 100 mM Tris 8.5, and digested with 0.5 mg ml⁻¹ of trypsin supplemented with CaCl₂ (at a final concentration of 1 mM) overnight at 37 °C. Trypsin-digested samples were submitted to the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley for analysis. Peptides were processed using multidimensional protein identification technology (MudPIT) and identified using a LTQ XL linear ion trap mass spectrometer. To identify high-confidence interactors, CompPASS analysis of the query mass spectrometry result was performed against mass spectrometry results from unrelated Flag immunoprecipitates performed in our laboratory.

For TMT labelling, samples were prepared in the same manner as previously described³⁷. Following trypsin digestion, however, samples were desalted using a C18 column (Agilent, A57203), dried overnight, resuspended in 80 µl of 200 mM HEPES, pH 8.0 and quantified using the Pierce Quantitative Colorimetric Peptide Assay kit (Pierce, 23275) on a microplate reader. Peptides were then normalized to equal masses in 100 µl volumes with 200 mM HEPES, pH 8. TMT labelling was performed using the TMTsixplex Isobaric Mass Tagging Kit (Thermo Fisher, 90066) per the manufacturer's instruction. Labelled peptides were combined in equal volumes, desalted, dried and identified using a Fusion Lumos mass spectrometer by the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory.

Immunofluorescence microscopy

For immunofluorescence analysis of neural inductions, H1 cells and H1 cells undergoing neural conversion were seeded on Matrigel-coated 96-well plates in mTeSR1 or STEMdiff neural induction medium plus 10 µM Y-27632 for 24 h, washed with 1× PBS plus 1 mM MgCl₂ and 1 mM CaCl₂, fixed with 4% paraformaldehyde buffered in 1× PBS for 15 min, permeabilized in 1× PBS plus 0.1% Triton X-100 for 10 min, blocked in 10% FBS plus 0.1% Triton X-100 for 30 min and stained with indicated antibodies diluted in 10% FBS plus 0.1% Triton X-100. Images were taken on an Opera Phenix High-Content Screening System (PerkinElmer) using a 40× air objective and processed using Harmony High Content Imaging and Analysis Software (PerkinElmer).

Live-cell imaging

H2B-mCherry expressing H1 cells were transfected with indicated siRNAs and seeded on Matrigel-coated 8-chamber microscopy slides

(Lab-Tell, 155409). Twenty-four to forty-eight hours after transfection, cells were imaged every 3 min for 12–14 h using a Zeiss LSM 710 confocal microscope with 20× objective. Mitotic cells were identified manually.

Analysis of cell-cycle progression

For DNA content analysis, single-cell suspensions were generated with trypsin, fixed for 15 min with 4% paraformaldehyde buffered in 1× PBS, washed with 1× PBS and incubated with 2 µg ml⁻¹ of Hoescht 33342 buffered in 1× PBS for 30 min at room temperature with gentle rocking. Single cells were filtered through a mesh strainer and analysed using an LSRFortessa flow cytometer (Becton Dickinson). Cytometry data were processed using the FlowCytometryTools Python package and custom-built Python scripts.

Sonication and ChIP-qPCR analysis

Cells were resuspended in 1× PBS and fixed at room temperature with 1% formaldehyde (Fisher, UN1198) for 10 min or with 1.5 mM ethylene glycol bis(succinimidyl succinate) (EGS) for 20 min followed by 1% formaldehyde for an additional 10 min. Crosslinking reactions were quenched with 125 mM glycine buffered in 1× PBS for 10 min. Crosslinked cells were washed twice with 1× PBS, collected, snap-frozen and stored at –80 °C for later use. Collected pellets were resuspended in sonication buffer (50 mM Tris 8.0, 10 mM EDTA, 1% SDS and 1× cOmplete protease inhibitor cocktail), incubated on ice for 15 min and pelleted at 2,000g. Pellets were washed 4× with sonication buffer and sonicated in 12 × 24-mm tubes (Covaris, 520056) at 150 W (peak power) using an S220 ultrasonicator (Covaris) with a duty factor of 20 and 200 cycles per burst for 30–35 cycles (30 s on and 30 s off). Sonicated extracts were clarified by high-speed centrifugation, snap-frozen and stored at –80 °C for later use. ChIP extracts were diluted 10-fold in dilution buffer (20 mM Tris 8.0, 167 mM NaCl, 1 mM EDTA, 1% Triton X-100 and 1× cOmplete protease inhibitor cocktail), precleared with protein G/A-agarose resin and bound overnight to the indicated antibodies (Supplementary Table 2) at 4 °C. Antibodies were pulled down by addition of BSA-blocked protein G/A-agarose resin. Beads were washed twice with low salt wash buffer (20 mM Tris 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), twice with high salt wash buffer (20 mM Tris 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), once with LiCl buffer (20 mM Tris 8.0, 250 mM LiCl, 1 mM EDTA, 1% deoxycholate and 1% Nonidet P-40) and twice with 1× TE. Samples were eluted twice at 30 °C with 1% SDS buffered in 1× TE. Eluates were pooled, treated with RNase A and reverse-crosslinked overnight at 65 °C. Samples were then treated with proteinase K, phenol:chloroform extracted, isopropanol precipitated and eluted in 10 mM Tris 8. Resuspended samples were quantified using the KAPA SYBR FAST Universal kit (Kapa Biosystems, KK406) on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). ChIP–qPCR primers used in this study can be found in Supplementary Table 3.

Real-time qPCR analysis

For real-time qPCR analysis, total RNA was purified from cells using the NucleoSpin RNA kit (Macherey-Nagel, no. 740955) or with acid phenol and reverse-transcribed using the Maxima First Strand cDNA Synthesis kit (Thermo Fisher, K1671). Expression levels were quantified using the Luna Universal qPCR Master Mix (NEB, M3003) on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). Real-time qPCR primers used in this study can be found in Supplementary Table 3.

Sonication and ChIP-seq analysis

For sonication and ChIP-seq analysis, samples were prepared as described in 'Sonication and ChIP–qPCR analysis'. Libraries were constructed by the Functional Genomics Laboratory at UC Berkeley, multiplexed and sequenced by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley on a HiSeq2500 or a HiSeq4000 (Illumina). Alignments for the paired-end or single-read sequencing runs were

Article

performed against the hg19 reference genome using Bowtie2. ChIP peaks were called with MACS14. Downstream analyses were performed using Bedtools and Deepools.

MNase ChIP-seq sample preparation

For MNase ChIP-seq analysis, human ES cells were collected by accurate treatment, washed once with ice-cold 1× PBS and resuspended in 1 ml of 1× PBS. Single-cell suspensions were crosslinked with 1% formaldehyde for 10 min at room temperature, quenched with glycine (at a final concentration of 125 mM) for 2 min, washed with 1× PBS, snap-frozen in liquid nitrogen and stored at −80 °C for later use. Frozen pellets were resuspended in an equal pellet volume of lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris 8.0, 1× cOmplete protease inhibitor cocktail and 1× phosSTOP), incubated on ice for 10 min, diluted 4-fold with dilution buffer (1% Triton X-100, 150 mM NaCl, 20 mM Tris 8.0, 2.5 mM CaCl₂, 1× cOmplete protease inhibitor cocktail and 1× phosSTOP), digested with 150 units of MNase (Worthington, LS004798) per 200 µl of pellet volume for 5 min at 37 °C, quenched with 6 mM EDTA and 6 mM EGTA, spun at 20,000g to remove debris, aliquoted, snap-frozen in liquid nitrogen and stored at −80 °C for later use. MNase-digested chromatin was precleared with protein-G dynabeads (Thermo, 10003D) and bound to indicated antibodies overnight at 4 °C. Antibodies were immunoprecipitated by addition of BSA-blocked protein-G dynabeads. Beads were washed twice with low salt wash buffer (20 mM Tris 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), twice with high salt wash buffer (20 mM Tris 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), once with LiCl buffer (20 mM Tris 8.0, 250 mM LiCl, 1 mM EDTA, 1% deoxycholate and 1% Nonidet P-40), and twice with 1× TE. Samples were eluted twice at 30 °C with 1% SDS buffered in 1× TE. Eluates were pooled, treated with RNase A and reverse-crosslinked overnight at 65 °C. Samples were then treated with proteinase K, phenol:chloroform extracted, isopropanol precipitated and eluted in 10 mM Tris 8.

MNase ChIP-seq library construction

Purified DNA (see ‘MNase ChIP-seq sample preparation’) was quantified using a Fragment Analyzer (Agilent). Twenty-five nanograms of purified DNA was resuspended up to 50 µl in water. Ten microlitres of T4 DNA ligase buffer (NEB, B0202), 4 µl of 10 mM dNTPs, 5 µl of T4 DNA polymerase (NEB, M0203), 1 µl of Klenow DNA polymerase (NEB, M0210), 5 µl of T4 DNA polynucleotide kinase (NEB, M0201) and 25 µl of water were added to the diluted input DNA and incubated at 25 °C for 30 min. Samples were purified with Ampure XP beads (Beckman, A36881) and resuspended in 32 µl of water. Five microlitres of buffer 2 (NEB, B7002), 1 µl of 10 mM dATP, 3 µl of Klenow fragment (NEB, M0212) and 9 µl of water were added to the end-repaired DNA and incubated at 37 °C for 30 min. Samples were purified with Ampure XP beads (Beckman, A36881) and resuspended in 23 µl of water. Five microlitres of Truseq Y adaptors for paired-end sequencing (custom-made), 5 µl of 10× ligase buffer (NEB, B0202), 1.5 µl of T4 DNA ligase (NEB, M0202) and 15.5 µl of water were added to the 3'-adenylated DNA and incubated at room temperature for 1 h. Samples were purified with Ampure XP beads (Beckman, A36881) and resuspended in 30 µl of water. Three microlitres of adaptor-ligated DNA was used for PCR amplification (KAPA HiFi master mix, KK201).

MNase ChIP-seq and analysis

MNase ChIP-seq samples (see ‘MNase ChIP-seq library construction’) were multiplexed and sequenced by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley on a HiSeq4000 (Illumina). Alignments for the single-read sequencing runs were performed against the hg19 reference genome using Bowtie2. ChIP peaks were called with MACS14. Downstream analyses were performed using Bedtools and Deepools.

RNA-sequencing sample preparation and analysis

Total RNA was purified from cells with TRIzol (Thermo, 15596026) and digested with TURBO DNase (Thermo, AM2238). Total RNA was poly(A)-selected and sequencing libraries were constructed using the KAPA mRNA HyperPrep kit (KK8580) by the Functional Genomics Laboratory at UC Berkeley. Libraries were sequenced by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley on a HiSeq4000 (Illumina). Gene-expression analysis was performed using Kallisto against hg19 as the reference genome.

Bioinformatics

Identified ChIP peaks were subjected to bioinformatic analyses. GO enrichment analyses were performed using DAVID 6.8 (<https://david.ncifcrf.gov>). Comparison with SAGE data was performed using the CGAP-SAGE feature of DAVID, a web-based application (<https://david.ncifcrf.gov>).

Purification of phosphomimetic APC/C–CDC20 with WDR5 for negative-stain electron microscopy

Recombinant APC/C–CDC20 containing glutamate mutations that mimic phosphorylation³⁸ was purified as previously described³⁹. In brief, APC/C and CDC20 were expressed independently in High Five insect cells (Thermo Fisher Scientific) and colysed by mixing and sonication. Cleared lysate was treated to tandem Strep- and GST-affinity chromatography selections for APC2 and APC16, respectively. GST elution fractions containing APC/C–CDC20 were combined with TEV protease, HRV14 3C protease and purified MBP–Flag–WDR5–His containing a TEV proteolytic site N-terminal of the Flag tag. This mixture was further purified through Flag affinity chromatography and eluted with antigenic peptides.

Negative-stain electron microscopy

For negative-stain electron-microscopy studies, 125 µg of purified APC/C–CDC20–WDR5 eluate from Flag immunoprecipitations was loaded onto a 10–40% glycerol gradient containing 50 mM HEPES pH 8.0, 200 mM NaCl and 2 mM MgCl₂. For particle fixation by GraFix⁴⁰, the gradient also contained 0.025% and 0.1% glutaraldehyde in the lighter and denser glycerol solution, respectively, creating an additional glutaraldehyde gradient from top to bottom (0.025–0.1%). Centrifugation was performed at 34,000 rpm in a TH-660 rotor (Thermo Fisher Scientific) for 15 h at 0 °C and the solution was subsequently fractionated. APC/C particles were allowed to adsorb on a thin film of carbon, transferred onto a plasma-cleaned lacey grid (LC200-CU, Electron Microscopy Services) and then stained for 2 min with a 4% (w/v) uranyl formate solution. Micrographs were collected on a FEI Titan Halo at 300 KV with a Falcon 2 direct detector (FEI) (MPI of Biochemistry). The nominal magnification was 45,000×, resulting in an image pixel size of 2.37 Å per pixel on the object scale and data were collected in a defocus range of 1.5–3.5 µm. Particles were autopicked using Relion⁴¹. The contrast transfer function parameters were determined using CTFFIND4⁴². Using Relion, particles were extracted from micrographs and subjected to 2D classification. Inconsistent class averages were removed before 3D classification in Relion.

Structural modelling was performed using UCSF Chimera (1.13.1)⁴³. To identify electron microscopy density corresponding to WDR5, the electron microscopy reconstruction of APC/C–CDC20–WDR5 obtained from 3D classification in Relion was superimposed with a prior map from an APC/C–CDC20–substrate complex (EMDB-3385, ref. ⁴⁴) low-pass-filtered to a comparable resolution. Although the resolution precludes definitive structural modelling, it allowed approximate, global placement of the crystal structures of WDR5⁴⁵, along with the APC2 winged-helix box and APC11 RING domains^{39,46}, which are known to be mobile and to adopt distinct orientations when bound to different APC/C partner proteins.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All original data are available from the corresponding author on request. ChIP-seq and RNA-sequencing data have been deposited at the Gene Expression Omnibus, accession code GSE122298.

Code availability

Custom Python scripts are available from the corresponding author on request.

34. Hockemeyer, D. et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* **29**, 731–734 (2011).
35. Chambers, S. M. et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
36. Kampmann, M., Bassik, M. C. & Weissman, J. S. Functional genomics platform for pooled screening and generation of mammalian genetic interaction maps. *Nat. Protoc.* **9**, 1825–1847 (2014).
37. McGourty, C. A. et al. Regulation of the CUL3 ubiquitin ligase by a calcium-dependent co-adaptor. *Cell* **167**, 525–538 (2016).
38. Qiao, R. et al. Mechanism of APC/CCDC20 activation by mitotic phosphorylation. *Proc. Natl Acad. Sci. USA* **113**, E2570–E2578 (2016).
39. Brown, N. G. et al. RING E3 mechanism for ubiquitin ligation to a disordered substrate visualized for human anaphase-promoting complex. *Proc. Natl Acad. Sci. USA* **112**, 5272–5279 (2015).
40. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
41. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
42. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

43. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. Zhang, S. et al. Molecular mechanism of APC/C activation by mitotic phosphorylation. *Nature* **533**, 260–264 (2016).
45. Zhang, P., Lee, H., Brunzelle, J. S. & Couture, J. F. The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res.* **40**, 4237–4246 (2012).
46. Brown, N. G. et al. Mechanism of polyubiquitination by human anaphase-promoting complex: RING repurposing for ubiquitin chain assembly. *Mol. Cell* **56**, 246–260 (2014).
47. Tsankov, A. M. et al. Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344–349 (2015).

Acknowledgements We thank N. Ingolia, R. Tjian, B. Schulman, J. Schaletzky and all members of M.R.'s laboratory for advice, helpful discussions and comments on the manuscript; and M. Matsumoto and V. Dixit for generously supplying us with linkage-specific ubiquitin antibodies. E.O. was funded by the Jane Coffin Childs Memorial Fund for Medical Research and the Siebel Stem Cell Institute. K.G.M. was funded by the NIH F32 postdoctoral fellowship (F32GM120956). A.M. was funded by the American Italian Cancer foundation and the California Institute for Regenerative Medicine. M.R. is an Investigator of the Howard Hughes Medical Institute. This work was also funded by an NIH grant (R01GM083064) awarded to M.R.

Author contributions E.O. performed work with stem cells (including knockdowns and infections), and performed and analysed the ultracomplex screen, cytometry-based studies, microscopy-based studies, qPCR-based studies and ChIP-seq and RNA-sequencing experiments. K.G.M. performed in vitro assays. E.O., K.G.M., A.M. and D.D.C. prepared samples for mass spectrometry analyses, performed immunoprecipitation experiments and maintained cell culture. E.R.W. and J.R.P. performed the cryo-EM of APC/C–WDR5. M.K. helped to analyse the ultracomplex screen. N.G. and C.Y.Z. prepared recombinant histones for in vitro assays. E.O., K.G.M. and M.R. interpreted the data and wrote the manuscript.

Competing interests M.R. is a cofounder and consultant to Nurix Therapeutics, a biotechnology company working in the ubiquitin space.

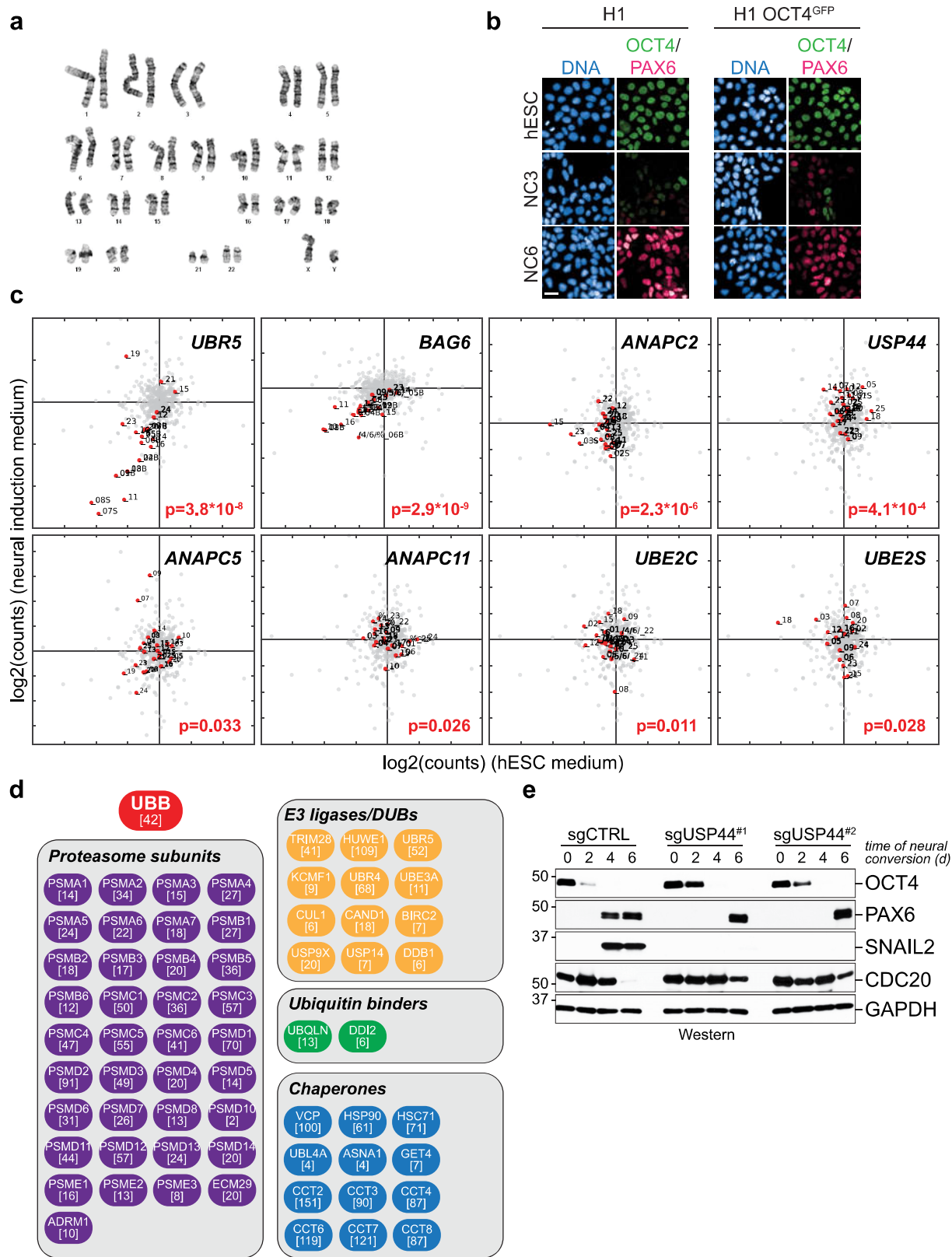
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2034-1>.

Correspondence and requests for materials should be addressed to M.R.

Peer review information *Nature* thanks William P. Tansey and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

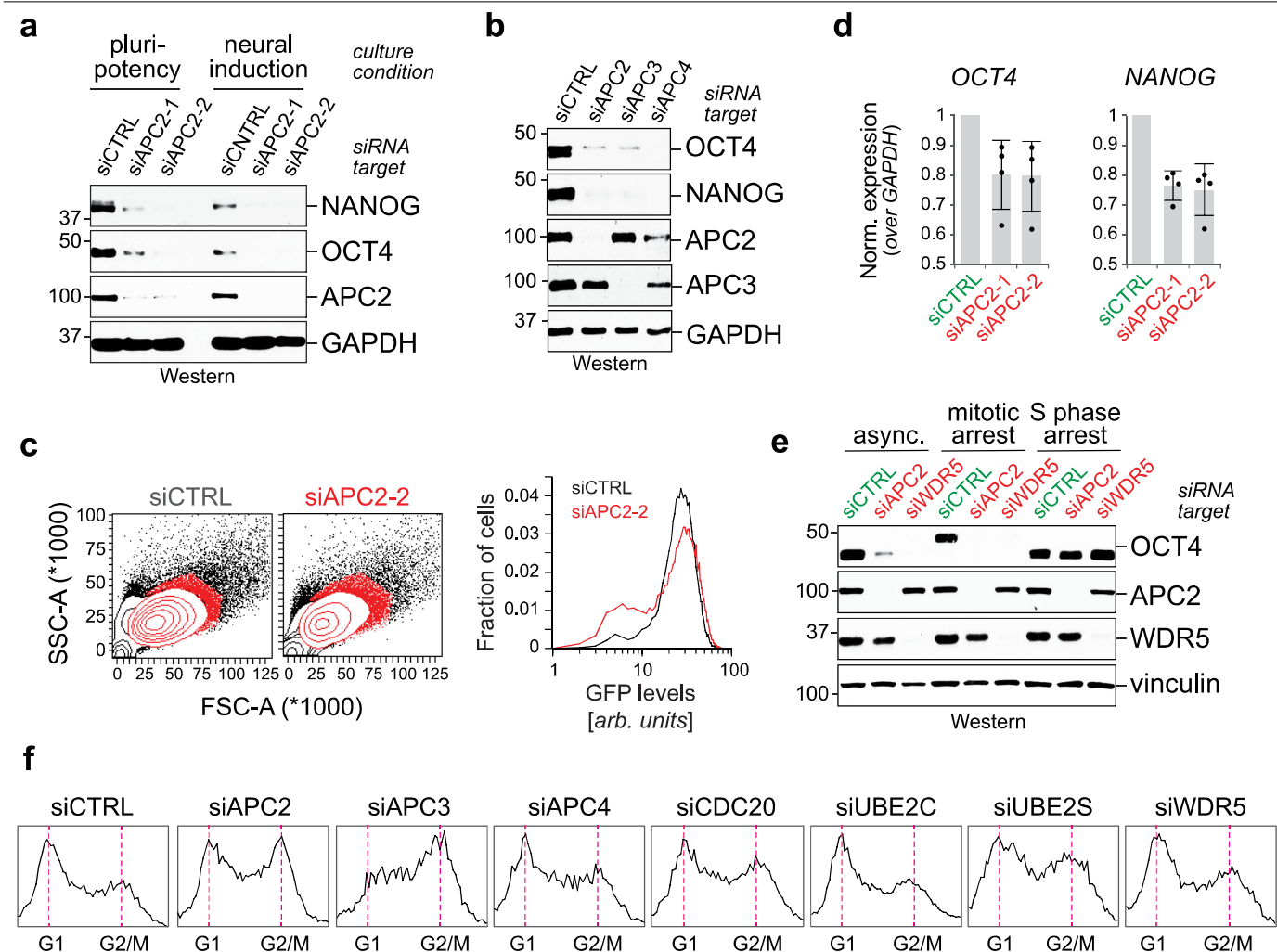
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | See next page for caption.

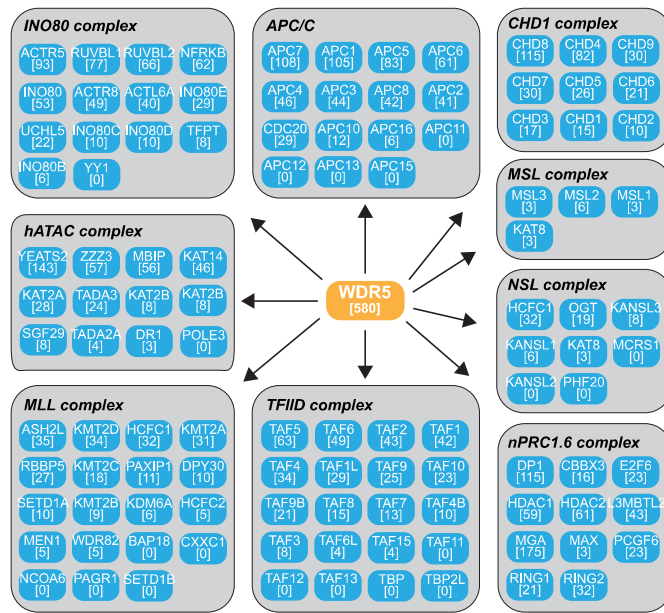
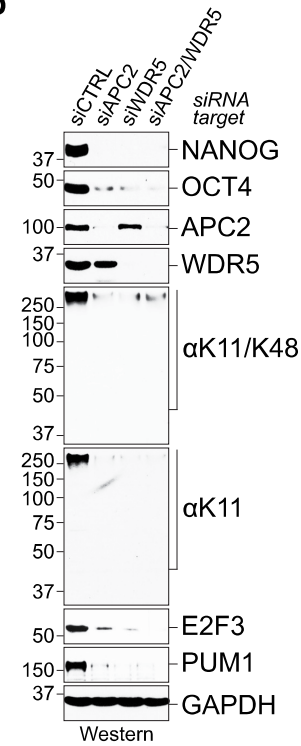
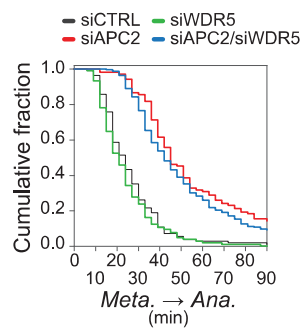
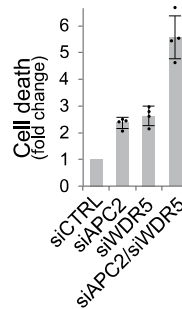
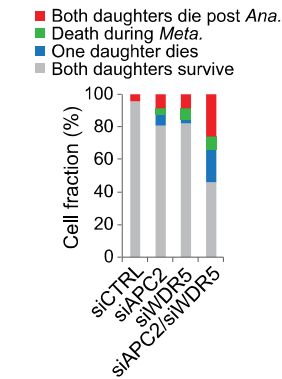
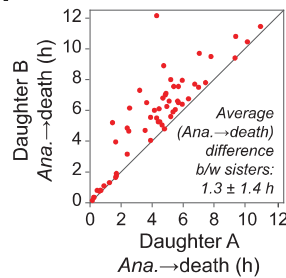
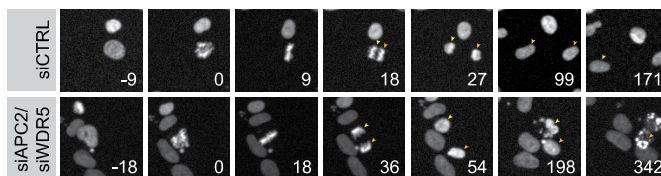
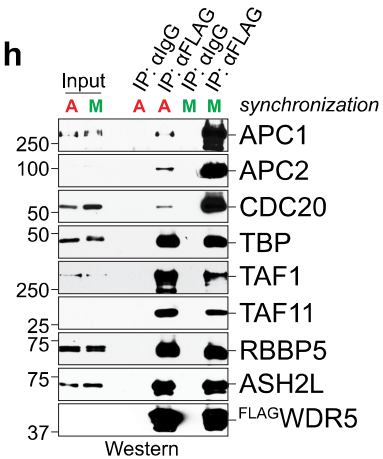
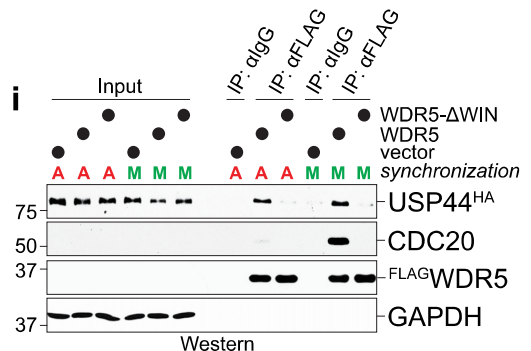
Extended Data Fig. 1 | Ultracomplex shRNA screen identifies APC/C and USP44 as regulators of human ES cell biology. **a**, Karyotype analysis of H1 OCT4–GFP cell line shows normal chromosome architecture. This line was karyotyped before performing the screen by a third party vendor (WiCell). Twenty cells were counted, 8 were analysed and 4 were karyotyped as normal. No clonal abnormalities were detected at the band resolution of 450–475. **b**, H1OCT4–GFP cells undergo neural conversion with an efficiency similar to that of the unmodified parent line. This experiment was performed three independent times with similar results. **c**, Deep sequencing read counts (log₂-transformed) for individual shRNAs (red dots) targeting the indicated gene from the screen in Fig. 1b. Grey dots represent negative-control shRNAs. *P* values (two-sided Mann–Whitney *U* test, not corrected for multiple

hypothesis testing) are indicated for each gene. **d**, Mass spectrometry analysis shows that many quality-control enzymes associate with K11/K48-branched chains in human ES cells. H1 human ES cells were synchronized in mitosis before being subjected to affinity purification using K11/K48-bispecific antibodies under denaturing conditions. Values listed in brackets are total spectral counts of tryptic peptides for each protein. **e**, CRISPR–Cas9-edited *USP44* H1 human ES cells show impaired rates of neural conversion. Expression of markers of pluripotency (OCT4), neural crest cells (SNAIL2) or neural progenitors (PAX6) were determined at indicated times of differentiation by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and western blotting using specific antibodies. This experiment was performed two independent times with similar results.



Extended Data Fig. 2 | Characterization of APC/C and the role of USP44 in pluripotency. **a**, Western blot of OCT4 and NANOG upon APC2 knockdown in asynchronous H1 human ES cells. This experiment was performed two independent times with similar results. **b**, Western blot of OCT4 and NANOG upon knockdown of APC/C subunits in asynchronous H1 human ES cells. This experiment was performed three independent times with similar results. **c**, Real-time qPCR of *OCT4* and *NANOG* upon APC2 knockdown in asynchronous H1 human ES cells (mean of $n = 4$ independent experiments, \pm s.d.). **d**, Flow cytometry analysis of APC2 depletion in H1 OCT4-GFP human ES cells. H1 OCT4-GFP human ES cells were transfected with siRNA against *APC2* for 48 h

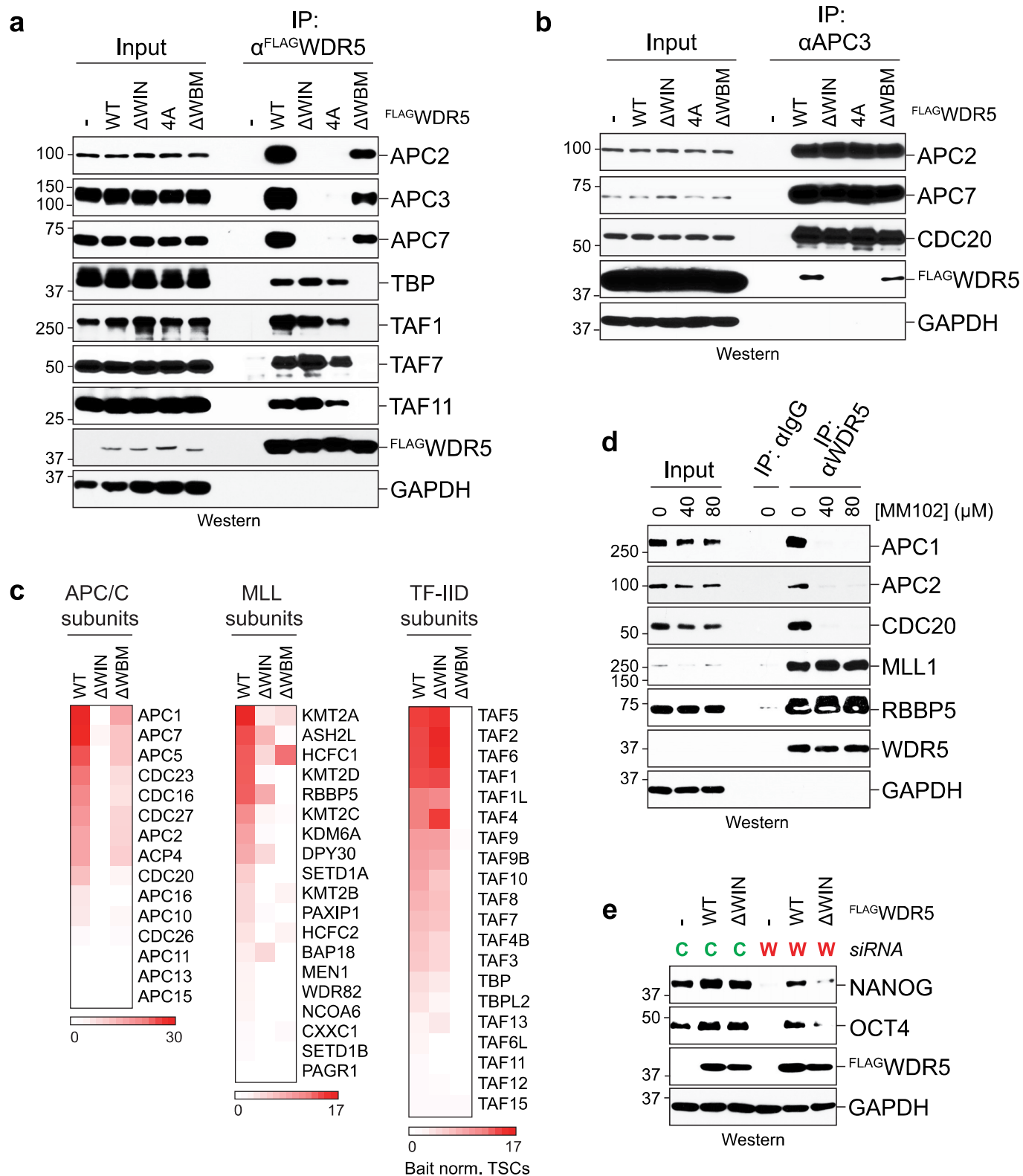
before cytometry analysis. This experiment was performed two independent times with similar results. **e**, Loss of the pluripotency marker OCT4 upon depletion of APC2 or WDR5 requires entry into mitosis. H1 human ES cells were transfected with indicated siRNAs for 36 h and treated with DMSO (asynchronous), 5 μ M STLC (mitotic arrest) or 200 mM thymidine (S-phase arrest) for an additional 12 h before collection for western blot analysis. This experiment was performed three independent times with similar results. **f**, Flow cytometry analysis of asynchronous H1 human ES cells transfected with indicated siRNAs for 72 h. This experiment was performed three independent times with similar results.

a**b****c****d****e****f****g****h****i****Extended Data Fig. 3** | See next page for caption.

Extended Data Fig. 3 | APC/C and WDR5 are required for human ES cell survival.

a, Mass spectrometry analysis of Flag-WDR5 purified from mitotic HEK293T cells. Values listed in brackets are total spectral counts of tryptic peptides of indicated proteins. **b**, Depletion of WDR5 phenocopies the depletion of APC2 in H1 human ES cells. H1 cells were depleted with the indicated siRNAs for 72 h before collection for western blot analysis. This experiment was performed once. **c**, A cumulative fraction curve measuring the length of each metaphase-to-anaphase transition. $n = 112$ cells for control siRNA; $n = 105$ cells for siRNA against *APC2*; $n = 106$ cells for siRNA against *WDR5*; and $n = 217$ cells for siRNAs against both *APC2* and *WDR5*. **d**, Depletion of APC2 or WDR5 causes cell death in H1 human ES cells. Cell death was measured by trypan blue staining of dead cells (mean of $n = 4$ independent experiments \pm s.d.). **e**, Quantifying cell survival using chromosome catastrophe as a proxy for cell death. H1 human ES cells virally expressing H2B-mCherry were transfected with the indicated siRNAs for 24 h before imaging by confocal microscopy. $n = 97$ cells for control siRNA; $n = 104$ cells for siRNA against *APC2*; $n = 90$ cells for siRNA against *WDR5*; and $n = 213$ cells for siRNAs against both *APC2* and *WDR5*. **f**, Sister cells die immediately following mitotic

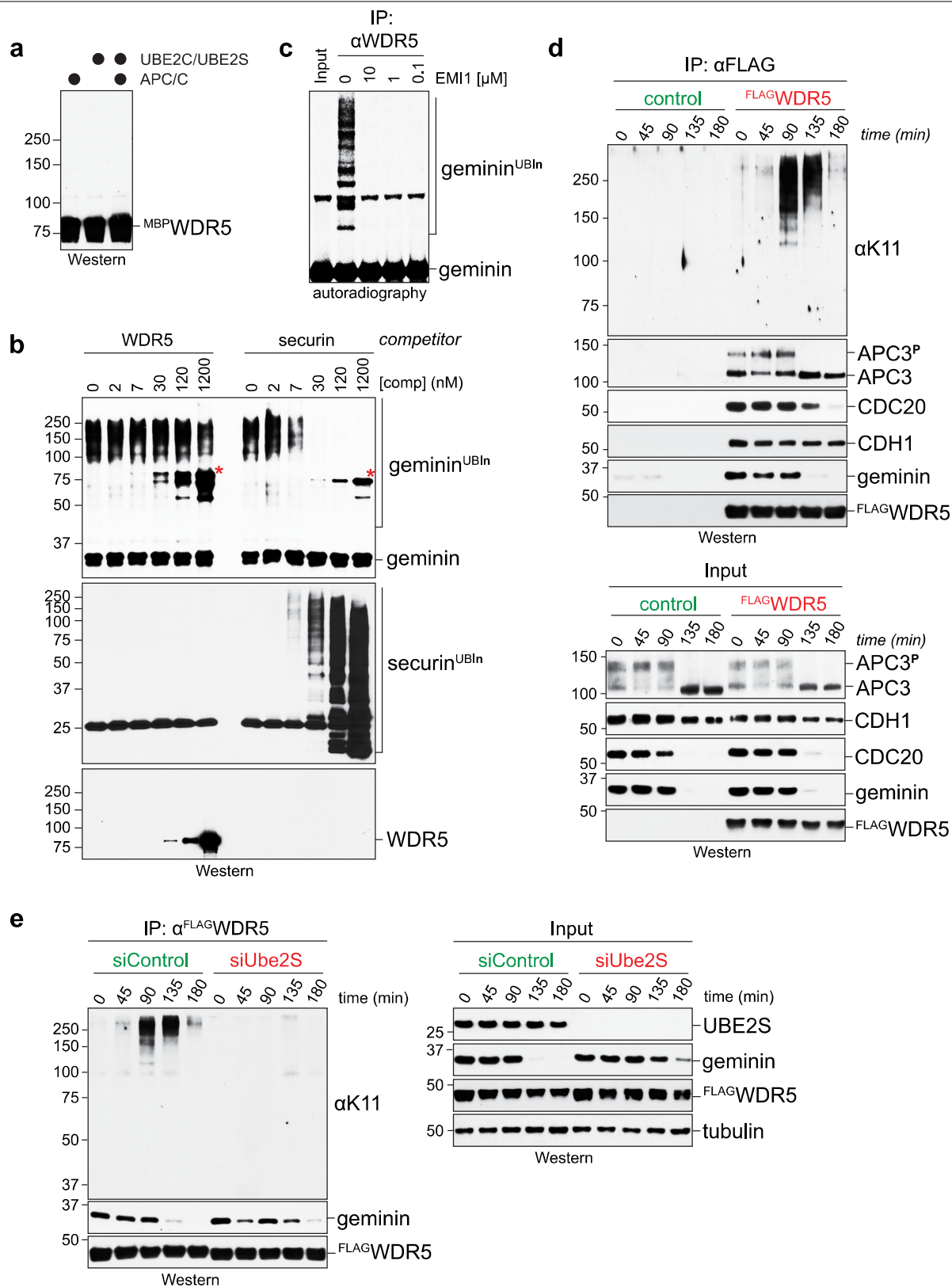
exit when depleted of APC2 and WDR5. H1 human ES cells virally expressing H2B-mCherry were transfected with siRNA against *APC2* and/or siRNA against *WDR5* for 24 h before imaging by confocal microscopy. The time of death, as defined by cells undergoing chromosome catastrophe, was measured for each sister (mean of $n = 57$ pairs of cells \pm s.d.). **g**, Representative frames of live-cell imaging from four independent experiments (in minutes) tracking the nuclei of siRNA-depleted H1 human ES cells virally expressing H2B-mCherry. Arrows mark individual sister cells upon mitotic exit. Chromosome catastrophe was used a proxy for cell death (time points 198 and 342). **h**, Flag-WDR5 associates with APC/C in mitotic H1 human ES cells. Flag-WDR5 immunoprecipitations were performed on asynchronous H1 human ES cells (A) or H1 human ES cells arrested in mitosis (M). Bound proteins were determined by SDS-PAGE and western blotting. This experiment was performed two independent times with similar results. **i**, Overexpressed haemagglutinin (HA)-tagged USP44 associates with Flag-WDR5 in both asynchronous and mitotic HEK293T cells. MYC-WDR5 was used as the control vector. This experiment was performed three independent times with similar results.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | WDR5 associates with APC/C and TBP on distinct surfaces. **a**, The WIN-motif binding site on WDR5 is critical for APC/C engagement, whereas the surface that binds the WBM is dispensable. A secondary binding surface (4A) is also important for the association of WDR5 with APC/C. The surface on WDR5 that binds the WBM is critical for TFIID association, whereas the WIN-motif binding site is dispensable. HEK293T cells were transfected with the indicated Flag-WDR5 variants, and cells were synchronized in mitosis. Flag-WDR5 was affinity-purified, and bound proteins were determined by western blotting. This experiment was performed five independent times with similar results. **b**, Reciprocal immunoprecipitations show that APC/C binds WDR5 through its WIN-motif binding site. Endogenous APC/C was purified from HEK293T cells expressing the indicated Flag-WDR5 variants, and bound proteins were determined by SDS-PAGE and western blotting. This experiment was performed three independent times with similar

results. **c**, Heat map of bait-normalized total spectral counts identified from Flag-WDR5-purified mass spectrometry experiments. HeLa cells were transfected with Flag-WDR5 for 24 h before mitotic synchronization. **d**, The WDR5 inhibitor MM-102 impairs the association of WDR5 with APC/C. Mitotic HeLa S3 cells were released into MM-102 for 2 h before immunoprecipitation experiments. Under these conditions, MM-102 did not prevent the association of WDR5 with MLL and RBBP5. This experiment was performed two independent times with similar results. **e**, Expression of wild-type WDR5 but not WDR5(Δ WIN) rescues the pluripotency defect caused by WDR5 depletion in H1 human ES cells. H1 human ES cells virally expressing siRNA-resistant WDR5 variants (WDR5 versus WDR5(Δ WIN)) were depleted of endogenous WDR5 (W) or treated with control siRNA (C). Expression of OCT4 and NANOG was determined by western blotting. This experiment was performed once.



Extended Data Fig. 6 | See next page for caption.

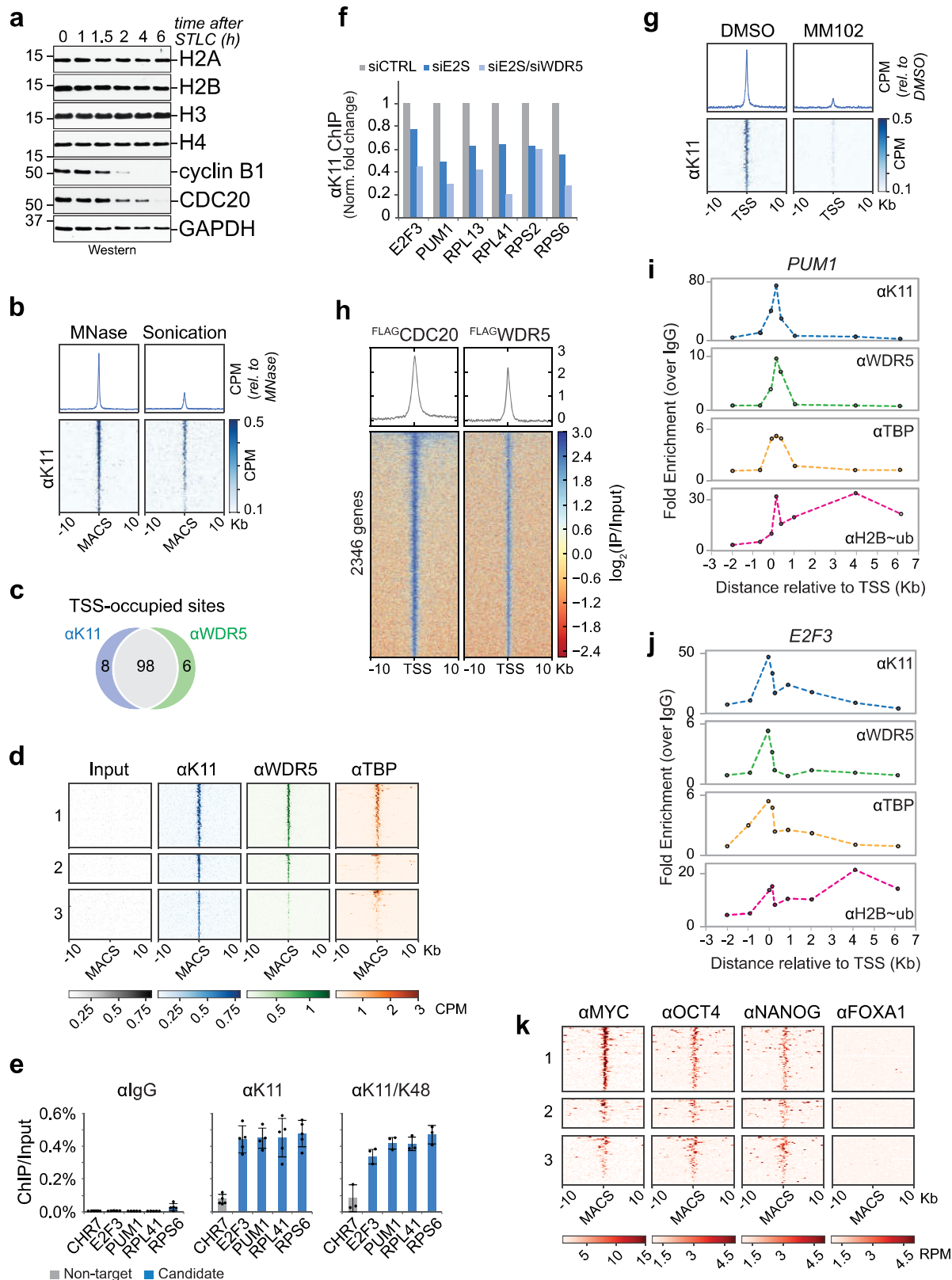
Extended Data Fig. 6 | WDR5 associates with active APC/C. **a**, APC/C does not ubiquitylate WDR5 in vitro. Recombinant WDR5 was incubated with active APC/C, E1, UBE2C, UBE2S and ubiquitin, and potential reaction products were detected by western blotting against WDR5. This experiment was performed once. **b**, APC/C-dependent ubiquitylation of geminin is outcompeted by recombinant securin (comp), a canonical substrate, but not by recombinant WDR5. Securin or WDR5 was added to APC/C-dependent geminin ubiquitylation reactions at the indicated concentrations, and various reaction products were detected using western blotting. Asterisks represent cross-reactive bands. This experiment was performed once. **c**, APC/C-WDR5-dependent ubiquitylation of geminin is inhibited by EMI1. WDR5 affinity

purifications from mitotic HeLa cells were incubated with E1, the APC/C-specific E2 enzymes UBE2C and UBE2S, and ubiquitin. EMI1 was added at indicated concentrations, and reaction products were detected by western blotting using antibodies against geminin. This experiment was performed two independent times with similar results. **d**, Immunoprecipitation of Flag-WDR5 from mitotic HEK293T cells coprecipitates K11-linked ubiquitin chains. HEK293T cells arrested in prometaphase were released into fresh medium, and WDR5 was affinity-purified at the indicated time points. Bound proteins were detected by western blotting. This experiment was performed once. **e**, Depletion of UBE2S eliminates WDR5-associated K11-linked ubiquitin chains in mitotic HEK293T cells. This experiment was performed once.

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Mitotic APC/C–WDR5 complexes are catalytically active. **a**, APC/C–WDR5 ubiquitylates human H3 in H3–H4 tetramers in vitro. APC/C was affinity-purified from mitotic HeLa cells and incubated with E1, UBE2C, UBE2S, ubiquitin and human H3–H4 tetramers, as indicated. Reaction products were detected by western blotting using antibodies against H3 and H4. This experiment was performed once. **b**, APC/C ubiquitylation of H2B in *X. laevis* H2A–H2B histone dimers. APC/C–WDR5 was purified from mitotic HeLa cells by Flag–WDR5 affinity purification and incubated with E1, UBE2C, UBE2S, ubiquitin and *X. laevis* histone octamers. Ubiquitylation was detected by western blotting against ubiquitylated H2B. This experiment was performed three independent times with similar results. **c**, APC/C–WDR5 ubiquitylation of H2B in *X. laevis* H2A–H2B–H3–H4 histone octamers. Reactions were performed as described in **b**. This experiment was performed two independent times with similar results. **d**, APC/C purified from H1 human ES cells is competent to ubiquitylate human H2B. This experiment was performed two independent times with similar results. **e**, APC/C purified from mitotic, but not S-phase, extracts can ubiquitylate H2B in vitro. APC/C was purified from HeLa cells synchronized at the indicated cell-cycle stages and incubated with E1, UBE2C, UBE2S, ubiquitin and *X. laevis* H2A–H2B dimers.

Histone ubiquitylation was detected by western blotting using antibodies against ubiquitylated H2B. This experiment was performed once. **f**, APC/C-dependent ubiquitylation of H2B requires K11 residue on ubiquitin for chain elongation. Ubiquitylation of H2A–H2B dimers by APC/C–WDR5 was performed as described in **e**, but with ubiquitin variants. This experiment was performed once. **g**, APC/C-dependent ubiquitylation of H2B requires both K11 and K48 on ubiquitin for synthesis of branched chains. This experiment was performed two independent times with similar results. **h**, Securin, a canonical APC/C substrate, outcompetes H2A–H2B dimers for APC/C-dependent ubiquitylation. The D-box motif (an APC/C–CDC20-specific degron) is required for full competition, whereas the KEN motif (an APC/C–CDH1-specific degron) is not. This experiment was performed two independent times with similar results. **i**, Polyubiquitylated H2B is degraded by the proteasome. K11/K48-branched chains were purified under denaturing conditions from mitotic HeLa cells either in the presence or absence of MG132, and modified H2B was detected using western blotting. Proteasome inhibition with MG132 was found to stabilize endogenous polyubiquitylated H2B. This experiment was performed four independent times with similar results.

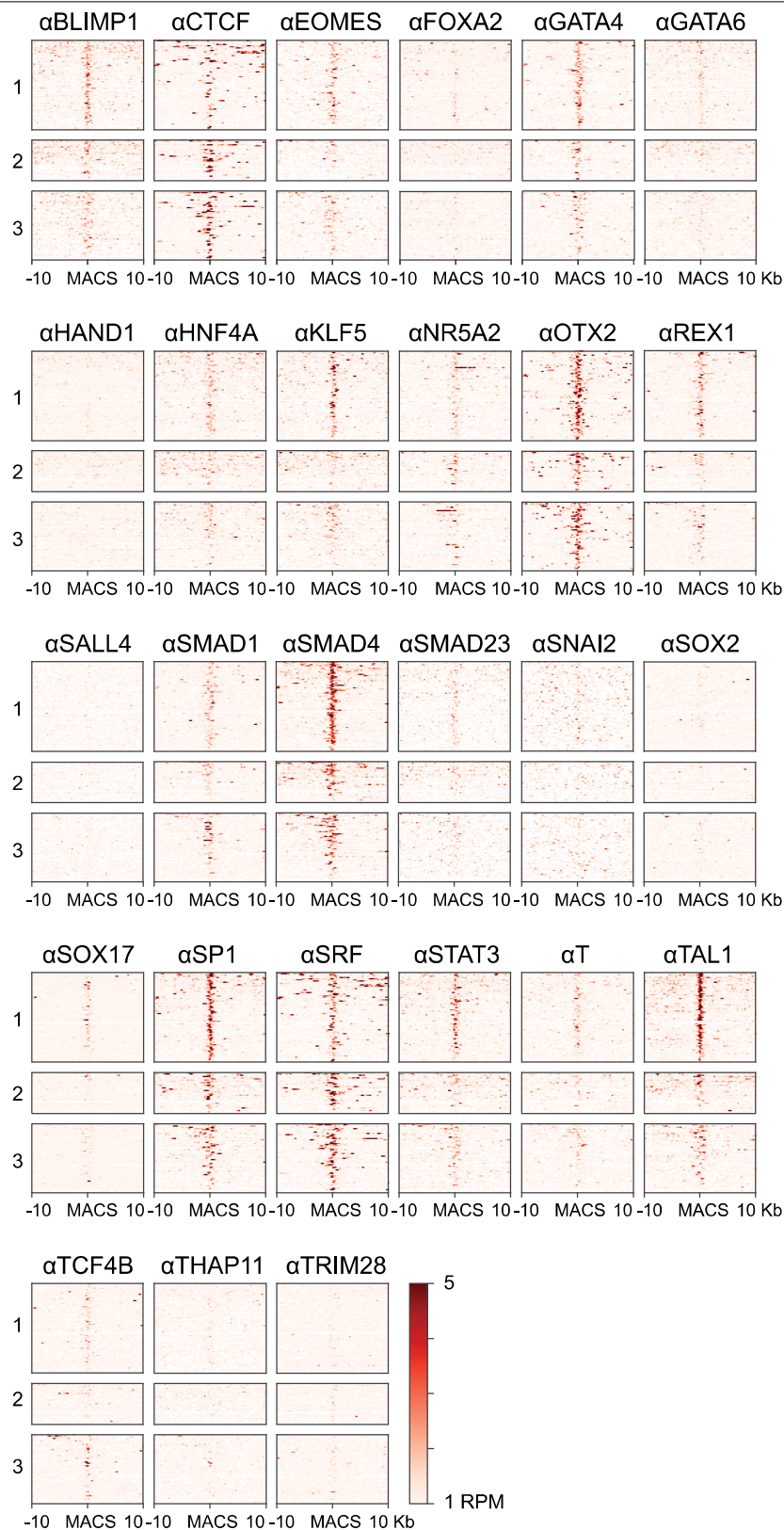


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | ChIP-seq analyses of APC/C- and WDR5-occupied

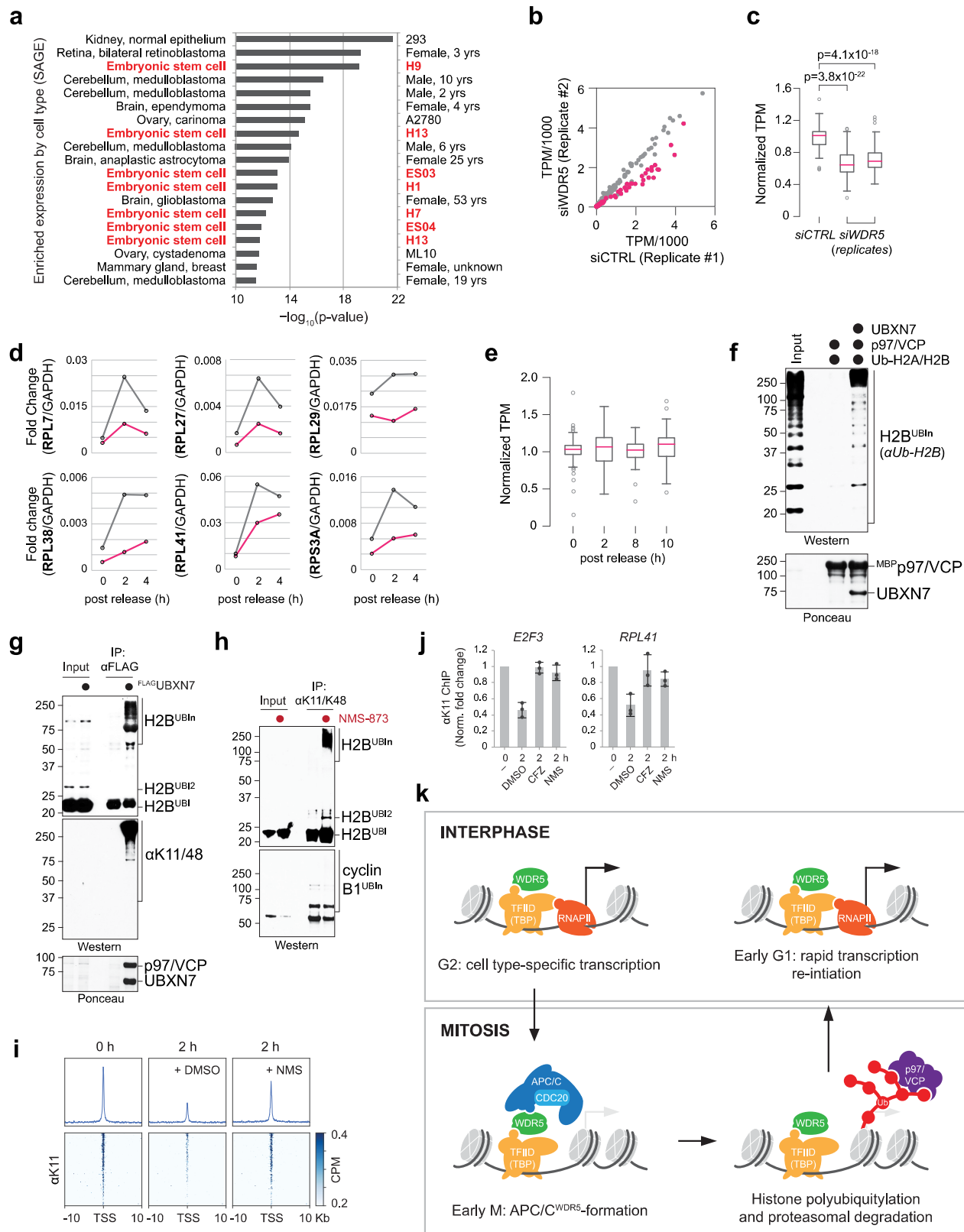
gene targets. a, Overall histone levels do not change upon release from mitosis. H1 cells were synchronized in mitosis by STLC and released into fresh medium. Indicated proteins were monitored by western blotting. This experiment was performed two independent times with similar results. **b,** Comparison between ChIP-seq and MNase ChIP-seq against anti-K11 from mitotic H1 human ES cells reveals that sonication shears polymeric ubiquitin linkages. **c,** Venn diagram of anti-K11 and anti-WDR5 ChIP peaks that colocalize with TSSs from MNase ChIP-seq experiments. MNase ChIP-seq experiments were performed from mitotic H1 human ES cells. **d,** Heat map of MNase ChIP-seq data from mitotic H1 human ES cells. Cluster 1 includes sites that are co-occupied by K11 and WDR5 near TSSs (within 100 bp); cluster 2 includes sites that are co-occupied by K11 and WDR5 outside of TSSs; and cluster 3 includes sites occupied only by K11, regardless of colocalization with TSSs. **e,** ChIP-qPCR analysis of candidate targets using K11- or K11/K48-linkage specific ubiquitin antibodies from mitotic H1 human ES cells. Mean of independent

replicates \pm s.d. $n = 3$ for K11/K48; $n = 5$ for IGG and K11 (except $n = 4$ for *PUM1*). **f,** ChIP-qPCR analysis of mitotic H1 human ES cells shows that K11 linkages synthesized at candidate sites are dependent on UBE2S and WDR5. This experiment was performed once. **g,** WDR5 inhibition prevents K11-ubiquitin chain formation at APC/C-WDR5-bound TSSs. H1 human ES cells were treated with or without 50 μ M MM102 during mitotic synchronization with STLC before anti-K11 MNase ChIP-seq. Heat map of all APC/C-WDR5-bound TSSs are shown. **h,** Heat map of ChIP-seq peaks of individual genes co-occupied by Flag-CDC20 and Flag-WDR5. ChIP-seq against anti-Flag was performed on mitotic HEK293T cells that overexpress Flag-CDC20 or Flag-WDR5. **i,** Spatial profile of *PUM1* of factor occupancy by ChIP-qPCR. This experiment was performed once. **j,** Spatial profile of *E2F3* of factor occupancy by ChIP-qPCR. This experiment was performed once. **k,** Heat map of MNase ChIP-seq data of transcription-factor binding. Previously published MNase ChIP-seq data were obtained⁴⁷, and APC/C-bound sites were analysed as described in **d**.



Extended Data Fig. 9 | Select transcription factors are found at APC/C–WDR5-bound sites. Heat map of MNase ChIP–seq data of transcription-factor binding. Previously published MNase ChIP–seq data were obtained⁴⁷, and APC/C-bound sites were analysed as follows: cluster 1 includes sites that are

co-occupied by K11 and WDR5 near TSSs (within 100 bp); cluster 2 includes sites that are co-occupied by K11 and WDR5 outside of TSSs; and cluster 3 includes sites only occupied by K11, regardless of colocalization with TSSs.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Regulation of chromatin and transcription by APC/C-WDR5.

a. Comparison of genes co-occupied by Flag-CDC20 and Flag-WDR5 from mitotic HEK293T cells with known gene-expression profiles reveals a strong overlap with ES cell and medulloblastoma cancer cell lines. $n=1,628$ genes were analysed (P values represent a one-sided Fisher's exact test with Bonferroni correction). **b.** Loss of APC/C-WDR5 function interferes with the expression of genes marked with K11-linked ubiquitin chains in H1 human ES cells. Poly(A)-selected RNA was purified from asynchronous H1 human ES cells transfected with control siRNA or siRNA against *WDR5* for 48 h, and subjected to RNA-sequencing analysis (a biological replicate of Fig. 4h). **c.** Transcript analysis of WDR5 depletion on APC/C-WDR5-dependent genes (from Fig. 4h and **b**). Box plots include the median TPM value ($n=90$ genes) with quartile ranges Q1-Q3; top whiskers represent the 3rd quartile + $1.5\times$ interquartile range; bottom whiskers represent the 1st quartile - $1.5\times$ interquartile range. P values were calculated from comparing individual TPM values of APC/C-WDR5-regulated genes ($n=90$) versus all transcripts ($n=18,791$) using a two-sided Student's t -test (unpaired). **d.** Real-time qPCR analysis of nascent RNA reveals APC/C-WDR5 target genes are reactivated upon mitotic exit and gene reactivation is dependent on WDR5. Initial screening from a single experiment. **e.** The RNA levels of genes regulated by APC/C-WDR5 do not change upon mitotic exit. RNA-sequencing analysis was performed on poly(A)-selected RNA purified from H1 human ES cells at the

indicated cell-cycle stages. Box plots were derived as described in **c** ($n=90$ genes). **f.** Ubiquitylated H2B preferentially associates with p97-UBXN7 in vitro. H2B was preubiquitylated by APC/C in vitro, and incubated with immobilized p97 or p97-UBXN7 complexes. Bound histone H2B was detected by western blotting. This experiment was performed three independent times with similar results. **g.** Flag-UBXN7 associates with polyubiquitylated H2B, p97 and K11/K48-linked branched ubiquitin chains in mitosis. Native Flag-UBXN7 immunoprecipitations were performed on mitotic HEK293T cells and bound proteins were detected by western blotting or Ponceau staining. This experiment was performed three independent times with similar results. **h.** H2B ubiquitylation is stabilized by p97 inhibition in cells. Denaturing K11/K48 immunoprecipitations were performed on H1 human ES cells synchronized in prometaphase or released into $10\mu\text{M}$ NMS-873 for 2 h. This experiment was performed four independent times with similar results. **i.** p97 inhibition restores K11 deposition at sites regulated by APC/C-WDR5 upon mitotic exit. Anti-K11 MNase ChIP-seq was performed from H1 human ES cells synchronized in mitosis (0 h) or released into fresh medium without (2 h + DMSO) or with p97 inhibition (2 h + $10\mu\text{M}$ NMS-873). **j.** Anti-K11 MNase ChIP-qPCR of candidate targets from mitotic H1 human ES cells (mean of $n=3$ independent replicates \pm s.d.). H1 human ES cells were synchronized in mitosis (0 h) and released into fresh medium for 2 h with the indicated drugs. **k.** Model of APC/C-dependent gene activation upon mitotic exit.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	ChIPseq data was mapped with Bowtie2 (2.3.3.1) and peaks were assigned with MACS14 (1.4.2). RNAseq data was mapped with Kallisto (0.44.0).
Data analysis	ChIPseq data was analyzed using Deeptools (2.5.3). RNAseq data was analyzed with Kallisto (0.44.0). GO analysis was performed with DAVID (6.8). Custom python scripts (2.7.15) were used to generate scatter and box plots (matplotlib package, 2.2.3). Flow cytometry analysis was performed using FlowCytometryTools (0.5.0), a python package (2.7.15). Harmony High Content Imaging and Analysis Software was used for microscopy analysis (4.9).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

ChIPseq and RNAseq data are banked at GEO (GSE122298)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No methods were used to predetermine sample size for experiments. For microscopy-based studies, we aimed to monitor at least 50 cells per condition. We determined this sample size based on the many years of microscopy-based cell cycle studies performed in our lab. For qPCR-based assays, no sample size calculation was performed, since hypotheses were validated using other metrics (ChIPseq).
Data exclusions	Data was excluded if cells failed to synchronize for experiments dependent on cell synchronization.
Replication	Biological replicates, if performed, have been clearly stated in the figure legends as independent experiments. For experiments without biological replicates, the hypothesis was tested using a different experimental setups to address the same question (i.e. reciprocal IPs, different techniques).
Randomization	No randomization was performed. Our experiments are not applicable for randomization since we did not have animal-based studies
Blinding	No experiments were blinded. Our experiments are not applicable for blinding since we did not have animal-based studies.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Anti-Actin, mouse (#69100, clone C4, MP Biomedicals, 1:1,000,000 for WB, Lot 04917), Anti-APC1, rabbit (#13329, clone D1E9D, Cell Signaling, 1:1000-2000 for WB, Lot 1), Anti-APC2, rabbit (#12301, Cell Signaling, 1:2000 for WB, Lot 1), Anti-APC3, mouse (#12530, clone D3I1V, Cell Signaling, 1:4000-5000 for WB, Lot D0814), Anti-APC3/CDC27, mouse (sc-9972, AF3.1, SCBT, 10 ug per IP, Lot F0717), Anti-APC4, rabbit (sc-20985, SCBT, 1:1000 for WB, Lot B0904), Anti-APC5, rabbit (sc-20986, SCBT, 1:1000 for WB, Lot G1003), Anti-APC6/CDC16, rabbit (sc-5615, SCBT, 1:1000 for WB, Lot H192), Anti-APC7, mouse (sc-365649, SCBT, 1:1000 for WB, Lot D1411), Anti-APC10, rabbit (sc-20989, SCBT, 1:1000 for WB, Lot A1604), Anti-APC11, rabbit (#14090, clone D1E7Q, Cell Signaling, 1:1000 for WB, Lot 1), Anti-ASH2L, rabbit (A300-489A, Bethyl Laboratories, 1:2000 for WB), Anti-CDC20, rabbit (#14866, clone D6C2Q, Cell Signaling, 1:4000-5000 for WB, Lot 1), Anti-CDH1, mouse (C7855, Sigma, 1:1000 for WB), Anti-Cyclin A, rabbit (sc-596, SCBT, 1:1000 for WB, Lot G1212), Anti-Cyclin B1, rabbit (#4138, Cell Signaling, 1:1000-2000 for WB, Lot 3), Anti-DYKDDDDK Tag, rabbit (#2368, Cell Signaling, 1:1000 for WB, Lot 12), Anti-DYKDDDDK Tag, rabbit (#14793, Cell Signaling, 1:10,000 for WB, Lot 4), Anti-E2F3, rabbit (GTX102302, GeneTex, 1:1000 for WB, Lot 39694), Anti-FLAG M2, mouse (F1804, SIGMA, 1 ug per ChIP, 10 ug per IP, 1:5000-100,000 for WB, Lot SLBW3851), Anti-GAPDH, rabbit (#5174, clone D16H11, Cell Signaling, 1:1,000,000 for WB, Lot 7), Anti-Geminin, rabbit (sc-13015, clone FL-209, SCBT, 1:2000 for WB, Lot D1610), Anti-HA Tag, rabbit (#3724, clone C29F4, Cell Signaling, 1:2000-10,000 for WB, Lot 9), Anti-H2A, rabbit (#12349, clone D603A, Cell Signaling, 1:1000-5000 for WB, Lot 1), Anti-H2B, mouse (#2934, Cell Signaling, 1:1000-100,000 for WB, Lot 4), Anti-ubiquitinyl-H2B (Lys120), rabbit (#5546, clone D11, Cell Signaling, 1-2 ug per ChIP, 1:1000 for WB, Lot 6), Anti-H3, rabbit (ab1791, Abcam, 1:1000-500,000 for WB, Lot GR3236370-1), Anti-H4, rabbit (ab10158, Abcam, 1:1000-10,000 for WB, Lot GR3186358-1), Anti-HSP90-β, rabbit (#7411, clone D3F2, Cell Signaling, 1:2000 for WB, Lot 1), Anti-K11 linkage-specific ubiquitin, human (Yau et al., 1-2 ug per ChIP, 10 ug per IP, 1:6000 for WB, Lab reagent), Anti-K11/K48 bispecific ubiquitin, human (Yau et al., 1-2 ug ChIP, 10 ug per IP, 1:300,000 for WB, Lab reagent), Anti-MLL1 carboxy-terminal, rabbit (#14197, clone D6G8N, Cell Signaling, 1:1000 for WB, Lot 1), Anti-Nanog, rabbit (#3580, Cell Signaling, 1:2000 for WB, Lot 3), Anti-OCT3/4, goat (sc-8628, SCBT, 1:500 for IF, Lot F2514), Anti-OCT3/4, rabbit (#2750, Cell Signaling, 1:2000 for WB, Lot 4), Anti-PAX6, rabbit (#901301, clone Poly19013,

BioLegend, 1:200 for IF, Lot B201255), Anti-PAX6, mouse (AB528427, DSHB, 1:2000 for WB), Anti-PUM1, rabbit (#12322, Cell Signaling, 1:1000 for WB, Lot 1), Anti-RBBP5, rabbit (#13171, clone D3I6P, Cell Signaling, 1:10,000 for WB, Lot 1), Anti-Securin, rabbit (sc-22772, clone H-160, SCBT, 1:2000 for WB, Lot D2204), Anti-SNAIL2, rabbit (#9585, Cell Signaling, 1:1000 for WB, Lot 6), Anti-TAF1, rabbit (#12781, clone D6J8B, Cell Signaling, 1:1000 for WB, Lot 1), Anti-TAF7, rabbit (#13506-1-AP, Proteintech, 1:1000 for WB), Anti-TBP, rabbit (#44059, clone D5C9H, Cell Signaling, 1 ug per ChIP, 1:10,000 for WB, Lot 1), Anti-Tubulin, mouse (#CP06, clone DM1A, EMD Millipore, 1:500,000 for WB, Lot 2681308), Anti-UBCH10/UBE2C, rabbit (A-650, Boston Biochem, 1:2000 for WB, Lot 2600774), Anti-UBE2S, rabbit (ab177508, Abcam, 1:10,000 for WB, Lot GR290784-7), Anti-Vinculin, rabbit (#4650, Cell Signaling, 1:1000 for WB, Lot 4), Anti-WDR5, mouse (sc-393080, clone G-9, SCBT, 10 ug per IP, Lot C2816), Anti-WDR5, rabbit (A302-430A, Bethyl Laboratories, 1:4000 for WB), Anti-WDR5, rabbit (#13105, Cell Signaling, 1:2000 for WB, Lot 1), Anti-WDR5, rabbit (C15410027, Diagenode, 1 ug per ChIP, Lot 001), Normal IgG, mouse (sc-2025, SCBT, 1-2 ug per ChIP, 1-10 ug per IP, Lot 33)

Validation

Antibodies validated by siRNA knockdown: Anti-APC2 (#12301), Anti-APC3 (#12530), Anti-CDC20 (#14866), Anti-UBCH10/UBE2C (A-650), Anti-UBE2S (ab177508), Anti-WDR5 (A302-430A), Anti-WDR5 (#13105), Anti-WDR5 (C15410027). Antibodies validated on recombinant proteins: Anti-DYKDDDDK (#2368), Anti-DYKDDDDK (#14793), Anti-FLAG M2 (F1804), Anti-HA (#3724), Anti-Geminin (sc-13015), Anti-H2A (#12349), Anti-H2B (#2934), Anti-ubiquitinyl-H2B(Lys120) (#5546), Anti-H3 (ab1791), Anti-H4 (ab10158), Anti-Securin (sc-22772). Antibodies validated by MS enrichment analysis: Anti-APC3/CDC27 (sc-9972), Anti-K11 (lab stock), Anti-K11/K48 (lab stock), Anti-WDR5 (sc-393080). Antibodies validated by the lab previously: Anti-Actin (#69100), Anti-APC1 (#13329), Anti-APC4 (sc-20985), Anti-APC5 (sc-20986), Anti-APC6/CDC16 (sc-5615), Anti-APC7 (sc-365649), Anti-APC10 (sc-20989), Anti-APC11 (#14090), Anti-CDH1 (C7855), Anti-Cyclin A (sc-596), Anti-Cyclin B1 (#4138), Anti-GAPDH (#5174), Anti-HSP90- β (#7411), Anti-Nanog (#3580), Anti-OCT3/4 (sc-8628), Anti-OCT3/4 (#2750), Anti-PAX6 (#901301), Anti-PAX6 (AB528427), Anti-SNAIL2 (#9585), Anti-Tubulin (#CP06), Anti-Vinculin (#4650). Antibodies validated from manufacturer's site: Anti-ASH2L (A300-489A, verified by heterologously transfecting ASH2L), Anti-E2F3 (GTX102302, verified by knockdown), Anti-MLL1 (#14197, verified by knockout), Anti-PUM1 (#12322, verified by Sur et al., 2018, Kulkarni et al., 2018), Anti-RBBP5 (#13171, verified by Bögershausen et al., 2015, Alvarado et al., 2017, Ishiushi et al., 2019), Anti-TAF1 (#12781, verified by Le Gallo et al., 2017), Anti-TAF7 (#13506-1-AP, uncharacterized), Anti-TBP (#44059, verified by Dal-Pra et al., 2017, Qin et al., 2018, Zhang et al., 2019).

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

H1 (WA01), HEK293T cells (Berkeley Cell Culture Facility), HeLa (lab stock), and Hela S3 (lab stock)

Authentication

H1 hESCs were purchased directly from WiCell (OCT4/NANOG positive, karyotype analysis showed no chromosomal anomalies). HEK293T were purchased directly from the Berkeley Cell Culture Facility (authenticated by short tandem repeat analysis). HeLa lines were not authenticated.

Mycoplasma contamination

All cell lines were routinely tested for mycoplasma contamination using the LONZA kit (MYCOALERT). All cell lines tested negative for mycoplasma.

Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified lines were used for this study.

ChIP-seq

Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122298>
(Pre-publication access token: mrgtoewspftcz)

Files in database submission

FlagCDC20_CHIP_rep1 (GSM3463651) FlagCDC20_Input_rep1 (GSM3463652) FlagWDR5-ChIP (GSM3463653) FlagWDR5-Input (GSM3463654) FlagCDC20_CHIP_rep2 (GSM3463655) FlagCDC20_Input_rep2 (GSM3463656) Sonication_ChIPseq_aK11_0h (GSM3844204) Sonication_ChIPseq_Input_0h (GSM3844205) MN_ChIPseq_Input_0h (GSM3844214) MN_ChIPseq_Input_2h (GSM3844215) MN_ChIPseq_Input_8h (GSM3844216) MN_ChIPseq_Input_10h (GSM3844217) MN_ChIPseq_K11_0h_rep1 (GSM3844218) MN_ChIPseq_K11_2h_rep1 (GSM3844219) MN_ChIPseq_K11_8h_rep1 (GSM3844220) MN_ChIPseq_K11_10h_rep1 (GSM3844221) MN_ChIPseq_TBP_0h (GSM3844222) MN_ChIPseq_TBP_2h (GSM3844223) MN_ChIPseq_TBP_8h (GSM3844224) MN_ChIPseq_TBP_10h (GSM3844225) MN_ChIPseq_K11_0h_rep2 (GSM3844226) MN_ChIPseq_K11_2h_rep2 (GSM3844227) MN_ChIPseq_K11_8h_rep2 (GSM3844228) MN_ChIPseq_K11_10h_rep2 (GSM3844229) MN_ChIPseq_WDR5_0h (GSM3844230) MN_ChIPseq_WDR5_2h (GSM3844231) MN_ChIPseq_WDR5_8h (GSM3844232) MN_ChIPseq_WDR5_10h (GSM3844233) MN_ChIPseq_Input_t0_DMSO (GSM3844238) MN_ChIPseq_Input_t0_MM102 (GSM3844239) MN_ChIPseq_Input_t2_DMSO (GSM3844240) MN_ChIPseq_Input_t2_NMS (GSM3844241) MN_ChIPseq_K11_t0_DMSO (GSM3844242) MN_ChIPseq_K11_t0_MM102 (GSM3844243) MN_ChIPseq_K11_t2_DMSO (GSM3844244) MN_ChIPseq_K11_t2_NMS (GSM3844245) MN_ChIPseq_Input_siCTRL (GSM3844246) MN_ChIPseq_Input_siCDC20 (GSM3844247) MN_ChIPseq_Input_siWDR5 (GSM3844248) MN_ChIPseq_K11_siCTRL (GSM3844249) MN_ChIPseq_K11_siCDC20 (GSM3844250) MN_ChIPseq_K11_siWDR5 (GSM3844251)

Genome browser session
(e.g. [UCSC](#))

Data submitted through GEO (see above).

Methodology

Replicates

Biological replicates for FLAG-CDC20 ChIPin 293Ts; technical replicates for aK11 mitotic release time course in H1s.

Sequencing depth

18-38 million reads per sample (paired-end sequencing) for ChIP performed in 293Ts, 12-90 million reads per sample (single read) for ChIP performed in H1s.

Antibodies

Flag M2 (Sigma, 1804), histone H3 (Abcam, 1791), K11 linked ubiquitin chains (lab reagent), TBP (Cell Signaling, 44059), WDR5 (Diagenode, C15410027)

Peak calling parameters

macs14 -t IP_sorted.bam -c Input_sorted.bam -n macs_filename -g 2.7e9 -S -w

Data quality

Quality metrics for ChIP of 293T cells given by macs14 (1.4.2): 99% (13835/14000) Flag-CDC20 peaks at FDR 5% and above 5-fold enrichment to input; 99% (4049/4090) Flag-WDR5 peaks at FDR 5% and above 5-fold enrichment to input. Quality metrics for ChIP of H1s performed for K11 and WDR5 using macs14 (1.4.2): Peaks were chosen with a 5% FDR cutoff, above a 10 to 12-fold enrichment compared to input, and a $-10 \times \text{LOG}_{10}(\text{pvalue})$ value greater than 100-120.

Software

MacS14 (1.4.2)

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Single cell H1 Oct4-GFP suspensions were treated with accutase and immediately analyzed.

Instrument

LSR Fortessa

Software

BD FACSdiva (6.2), FlowCytometryTools (0.5.0)

Cell population abundance

Measured Oct4 levels of all cells that passed FSC/SSC gates. Measured DNA content of cells that passed FSC/SSC gates.

Gating strategy

FSC/SSC gates to exclude cell debris.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Selective loading and processing of prespacers for precise CRISPR adaptation

<https://doi.org/10.1038/s41586-020-2018-1>

Received: 28 May 2019

Accepted: 19 December 2019

Published online: 19 February 2020

 Check for updates

Sungchul Kim^{1,5}✉, Luuk Loeff^{1,4,5}, Sabina Colombo¹, Slobodan Jergic^{2,3}, Stan J. J. Brouns¹ & Chirlmin Joo¹✉

CRISPR–Cas immunity protects prokaryotes against invading genetic elements¹. It uses the highly conserved Cas1–Cas2 complex to establish inheritable memory (spacers)^{2–5}. How Cas1–Cas2 acquires spacers from foreign DNA fragments (prespacers) and integrates them into the CRISPR locus in the correct orientation is unclear^{6,7}. Here, using the high spatiotemporal resolution of single-molecule fluorescence, we show that Cas1–Cas2 selects precursors of prespacers from DNA in various forms—including single-stranded DNA and partial duplexes—in a manner that depends on both the length of the DNA strand and the presence of a protospacer adjacent motif (PAM) sequence. We also identify DnaQ exonucleases as enzymes that process the Cas1–Cas2-loaded prespacer precursors into mature prespacers of a suitable size for integration. Cas1–Cas2 protects the PAM sequence from maturation, which results in the production of asymmetrically trimmed prespacers and the subsequent integration of spacers in the correct orientation. Our results demonstrate the kinetic coordination of prespacer precursor selection and PAM trimming, providing insight into the mechanisms that underlie the integration of functional spacers in the CRISPR loci.

CRISPR (clustered regularly interspaced short palindromic repeats) and Cas (CRISPR-associated) proteins constitute an RNA-guided adaptive immune system that defends prokaryotes against invading nucleic acids¹. The first step of immunity is adaptation, or spacer acquisition, in which the genetic memory is updated from foreign DNA fragments (prespacers) through the integration of spacers between the repeats of the CRISPR loci^{8,9}. Adaptation relies on the highly conserved Cas1 and Cas2 proteins, which form a heterohexameric Cas1₍₄₎–Cas2₍₂₎ integrase complex (Cas1–Cas2)^{2–5}. In vivo studies have suggested that Cas1–Cas2 identifies suitable prespacers on the basis of a PAM sequence, which is another prerequisite for CRISPR interference in immunity^{10,11}.

Profiling of spacers showed that Cas1–Cas2 derives new spacers from degradation intermediates that are produced by RecBCD during the repair of double-stranded DNA breaks¹² and by Cas3 during primed adaptation mediated by the Cascade–Cas3 complex^{13–15}. Notably, both RecBCD and Cas3 generate single-stranded DNA (ssDNA) degradation products that contrast with the optimal substrate for integration by Cas1–Cas2^{13,16–18}. For example, the integration-competent prespacer DNA for *Escherichia coli* type I-E Cas1–Cas2 is composed of a central 23-base-pair (bp) duplex with two 5-nucleotide (nt) single-stranded 3′-overhangs (here termed the canonical prespacer)^{2,3,5}. The ssDNA fragments that are generated by RecBCD and Cas3 must be re-annealed and processed into the canonical form before integration into the CRISPR locus. However, whether Cas1–Cas2 has an active role in capturing ssDNA fragments from these sources and making an effective Cas1–Cas2–prespacer integration complex remains unknown.

Spacers must ultimately be integrated in the correct orientation with respect to the position of the PAM sequence in the 3′-overhang^{2,4}. Cas1–Cas2 has been shown to integrate spacers in either orientation with an equal probability in vitro^{19–22}. However, only correctly oriented spacers result in functional CRISPR RNAs for target recognition in vivo^{21,22}. Recent in vivo and in vitro studies in type I-A, I-B, I-C and I-D CRISPR–Cas systems have shown that Cas4 has a critical role in the maturation of prespacer DNA and high-fidelity spacer integration^{23–27}. However, how systems that lack Cas4—such as *E. coli* type I-E—coordinate the correct orientation of new spacers is unknown.

Prespacer precursor selection by Cas1–Cas2

Cas1–Cas2 repurposes DNA fragments from invading genetic elements^{12,13}. Most of these DNA fragments will have structures that deviate from the canonical form of prespacer DNAs⁵. Cas1–Cas2 should therefore be able to bind non-canonical prespacer DNAs (precursor prespacer DNAs). To visualize the process of precursor prespacer loading at a high spatiotemporal resolution, we developed a single-molecule Förster resonance energy transfer (smFRET) assay (Fig. 1a, Extended Data Fig. 1a). In brief, biotinylated Cas1–Cas2 complexes were immobilized on a microscope slide through biotin–streptavidin linkage and presented with prespacer DNAs labelled with a donor (Cy3) and an acceptor (Cy5) dye on the top and bottom strands, respectively (Fig. 1b). These labelling positions yielded a FRET value of around 0.72 (Extended Data Fig. 1b) and enabled us to examine the binding events in

¹Kavli Institute of Nanoscience, Department of Bionanoscience, Delft University of Technology, Delft, The Netherlands. ²Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, New South Wales, Australia. ³Illawarra Health and Medical Research Institute, Wollongong, New South Wales, Australia. ⁴Present address: Department of Biochemistry, University of Zurich, Zurich, Switzerland. ⁵These authors contributed equally: Sungchul Kim, Luuk Loeff. ✉e-mail: sungchulkim.kr@gmail.com; c.joo@tudelft.nl

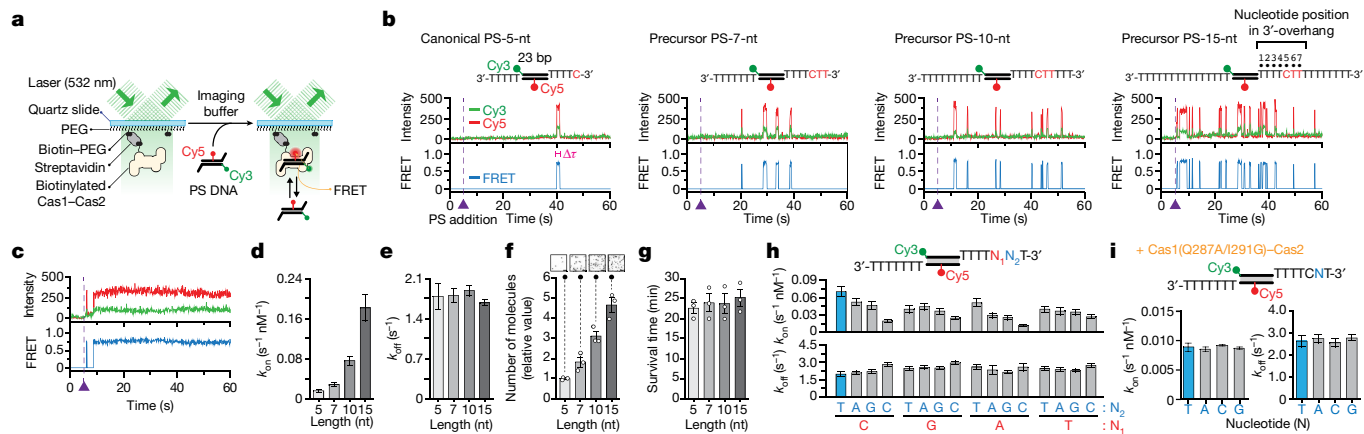


Fig. 1 | Single-molecule analysis of prespacer precursor selection by Cas1–Cas2. **a**, Single-molecule assay to probe Cas1–Cas2 binding to canonical and precursor prespacer (PS) DNA. PEG, polyethylene glycol. **b**, Time traces of donor (Cy3, green) and acceptor (Cy5, red) fluorescence signals and FRET efficiency (blue) exhibiting short-lived binding, where Δt is the dwell time of prespacer DNA-binding events to Cas1–Cas2. Canonical and precursor prespacer constructs consist of a 23-bp central duplex, a top strand Cy3-labelled at the 5′-end and a bottom strand Cy5-labelled at the 16th nucleotide from the 5′-end. DNA was added at $t = 5$ s. **c**, Time trace of long-lived binding. DNA was added at $t = 5$ s. **d**, Binding frequencies (k_{on}) calculated from the cumulative probability of the arrival time. **e**, Dissociation rates (k_{off}) of

short-lived events calculated from the dwell time of binding events by single-exponential fitting. **f**, Relative number of molecules at 30 min after addition of canonical or precursor prespacers ($n = 3$). Representative CCD images (acceptor channel) are included as insets. Scale bars, 5 μ m. **g**, Survival time of stably bound canonical or precursor prespacers on Cas1–Cas2. Free DNA was washed away after a 30-min incubation ($n = 3$). **h**, k_{on} and k_{off} of short-lived events for various sequences at the PAM position. **i**, k_{on} and k_{off} of PAM variants with Cas1(Q287A/I291G)–Cas2. Data are mean \pm s.e.m. from 3 independent replicates (**f**, **g**), or mean \pm 95% confidence interval (CI), obtained by bootstrap analysis of a single replicate with $n \geq 500$ individual molecules (**d**, **e**, **h**, **i**). All data are representative of three replicates with similar results.

real time using total internal reflection fluorescence (TIRF) microscopy (Fig. 1a, Extended Data Fig. 1a).

We assessed the properties of Cas1–Cas2 binding to a canonical prespacer DNA (PS-5-nt) (Fig. 1b) with an optimal form for spacer integration into the CRISPR loci^{2–5}. Notably, the majority of binding events (more than 99%) showed transient interactions (Fig. 1b, Extended Data Fig. 1c); by contrast, long stable-binding events were rarely observed (less than 1%) (Fig. 1c). Experiments in the absence of Cas1–Cas2 showed negligible non-specific binding (Extended Data Fig. 1d). Control experiments in which the position of the dye on the precursor prespacer DNA was varied did not show any substantial differences in binding kinetics (Extended Data Fig. 1e–i).

To examine the binding of Cas1–Cas2 to precursor prespacer DNAs, we tested a series of substrates with 3′-overhangs of more than 5 nt in length (7, 10 and 15 nt) (Fig. 1b). Analysis of these binding events showed that the binding frequency (k_{on}) increased with longer 3′-overhangs (Fig. 1d, Extended Data Fig. 1j), whereas the average dwell time (τ_{off}) (equal to $1/k_{off}$, in which k_{off} is the dissociation frequency) remained unchanged (Fig. 1e, Extended Data Fig. 1c). After further extension of the 3′-overhang (20, 30, 40, 70 and 90 nt; Extended Data Fig. 1k), k_{on} saturated at an overhang length of around 40 nt (Extended Data Fig. 1l), whereas k_{off} remained comparable among the different overhang lengths (Extended Data Fig. 1m). This suggests that the length at which Cas1–Cas2 makes effective interactions with prespacer precursors is limited by a distance of around 40 nt from the duplex region. As a result of the frequent interactions with precursor prespacer DNAs, the number of stably bound molecules increased proportionally to the length of the 3′-overhang (Fig. 1f), while the survival rate of the stably bound molecules remained constant (Fig. 1g, Extended Data Fig. 1n). Furthermore, Cas1–Cas2 selected precursor prespacer DNAs with long 3′-overhangs effectively in a competitive environment (Extended Data Fig. 1o–q). These results are consistent with the binding behaviour observed in electromobility shift assays (EMSAs) (Extended Data Fig. 1r) and previous studies^{5,28}.

Next, we used a single-molecule assay to investigate whether Cas1–Cas2 recognizes the PAM sequence during the short-lived interactions. First, we explored how the position of the PAM site in the 3′-overhang

affects the binding behaviour of Cas1–Cas2. When the PAM sequence in the 3′-overhang was moved away from the optimal position, fewer molecules were stably bound (Extended Data Fig. 2a, b), suggesting that Cas1–Cas2 probed the 3′-overhang for PAM recognition. This finding is consistent with previous studies that showed that Q287 and I291 in the proline-rich C-terminal tail of Cas1b recognize C and T in positions +5 and +6, respectively, of the 5′-C₊₅T₊₆T₊₇-3′ region of the PAM sequence (5′-CTT-3′ PAM) in the 3′-overhang⁵ (Extended Data Fig. 2c).

To further elucidate the PAM specificity of Cas1–Cas2, we generated 16 different precursor prespacer DNAs that encompass all of the nucleotide combinations at the first and second positions (positions +5 and +6) of the optimally located PAM. Cas1–Cas2 interacted with the different PAM variants at distinct frequencies (Fig. 1h, Extended Data Fig. 2d), while the dissociation rate remained unaltered (Fig. 1h). Among these substrates, Cas1–Cas2 showed the highest binding frequency for 5′-CTT-3′ PAM, suggesting that Cas1–Cas2 preferentially binds precursor prespacer DNAs that contain a PAM sequence. The difference in binding frequency among substrates was no longer observed in a C-terminal tail mutant (Q287A/I291G) of Cas1–Cas2 (Fig. 1i). In addition, Cas1–Cas2 selected PAM-containing prespacer precursors through frequent interactions in a competitive environment (Extended Data Fig. 2e–g). This is consistent with *in vivo* data showing that spacers are preferentially acquired from PAM-flanking sites^{29,30}.

Cas1–Cas2 facilitates ssDNA pairing

Studies have shown that Cas1–Cas2 repurposes DNA degradation products from RecBCD and Cas3^{6,12,13}. These fragments are likely to be single-stranded when being released from the enzymes¹⁴. To test whether Cas1–Cas2 captures two complementary ssDNAs to form an effector integration complex, we repeated the single-molecule assay with a series of ssDNA fragments. Although binding events were dictated by the presence of the PAM sequence (Extended Data Fig. 3a), binding frequencies (Fig. 2a, Extended Data Fig. 3b) and dissociation rates (Fig. 2b, Extended Data Fig. 3c) remained unaltered with an increasing number of PAM sites in the ssDNA fragments. By contrast, the binding frequency, but not the dissociation rate, depended on the length of the

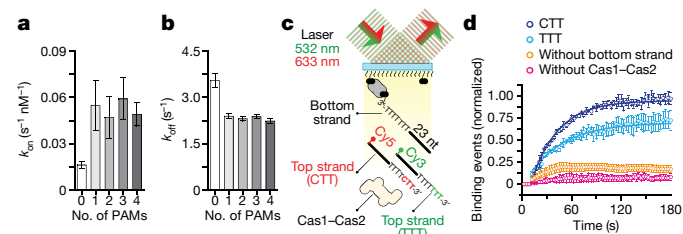


Fig. 2 | PAM-dependent ssDNA capture and facilitated pairing by Cas1–Cas2. **a**, **b**, k_{on} (**a**) and k_{off} (**b**) values of ssDNA with various numbers of PAM sequences binding to surface-immobilized Cas1–Cas2. **c**, Single-molecule assay for Cas1–Cas2-facilitated strand pairing. A biotinylated strand was surface-immobilized. The non-PAM (5'-TTT-3') strand was labelled with Cy3 and the PAM (5'-CTT-3') strand was labelled with Cy5. Green (532 nm) and red (640 nm) lasers were used to excite Cy3 and Cy5, respectively. **d**, Number of binding events over time. Solid lines represent a single-exponential fit. The k_{on} values ($0.031 \pm 0.003 \text{ s}^{-1} \text{ nM}^{-1}$ for CTT and $0.027 \pm 0.005 \text{ s}^{-1} \text{ nM}^{-1}$ for TTT) were calculated from the cumulative probability. The data were normalized from 0 to 1 by using the highest and lowest number from the CTT prespacer precursor. Data are mean \pm 95% CI, obtained by bootstrap analysis of a single replicate with $n \geq 500$ individual molecules (**a**, **b**). All data are representative of three replicates with similar results.

flanking sequence (Extended Data Fig. 3d, e). These findings suggest that Cas1–Cas2 might use a facilitated diffusion mechanism to locate PAM sequences on ssDNA (Extended Data Fig. 3f).

Given that prespacer DNAs need to be duplexed for spacer integration, we investigated whether Cas1–Cas2 facilitates the pairing between the PAM-containing top strand and its complementary bottom strand by designing a single-molecule DNA-capture assay (Fig. 2c). After immobilization of PAM-deficient bottom strands on the surface, we introduced two complementary top strands at 1 nM: a Cy5-labelled PAM-containing strand and a Cy3-labelled PAM-deficient strand (Fig. 2c). In the presence of Cas1–Cas2, the top strands accumulated rapidly; conversely, no detectable accumulation of binding was detected in the absence of either Cas1–Cas2 or the biotinylated bottom strand. Moreover, the PAM-containing top strand exhibited a higher binding affinity than the PAM-deficient version, resulting in a higher number of PAM-containing than PAM-deficient strands becoming annealed (Fig. 2d). These results are consistent with data from EMSA experiments (Extended Data Fig. 3g). Therefore, Cas1–Cas2 facilitates the pairing of complementary ssDNA fragments into precursor prespacer DNAs, with the PAM sequence acting as an identification marker.

Precursor prespacer trimming by DnaQ enzymes

Once the PAM-containing prespacer precursors with long 3'-overhangs are selected by Cas1–Cas2, the overhangs need to be processed into the canonical size of 5 nt for efficient integration into the CRISPR locus⁴ (Extended Data Fig. 4). Although a previous study suggested that Cas1–Cas2 itself could process the 3'-overhang into the optimal 5-nt size using the potential endonucleolytic activity of Cas1 subunits⁵, we were not able to reproduce this activity (Fig. 3a, Extended Data Fig. 5a). This suggests that there may be an alternative mechanism to produce canonical prespacer DNAs.

In the *Streptococcus thermophilus* type I-E CRISPR–Cas system, Cas2 is fused with a DnaQ-like domain that exhibits prespacer maturation activity³¹. Thus, we hypothesized that 3'-5' exonucleases with DnaQ-like domains might naturally act as 3'-overhang-trimming enzymes ('trimmers') in *E. coli*. To test whether any 3'-5' exonucleases with or without DnaQ-like domains could trim prespacer precursors into the mature versions, we tested the following enzymes: DNA polymerase I (Dna PolI), which contains an exonuclease II (ExoII) motif; the core complex (α , ϵ and θ subunits) of DNA polymerase III holoenzyme

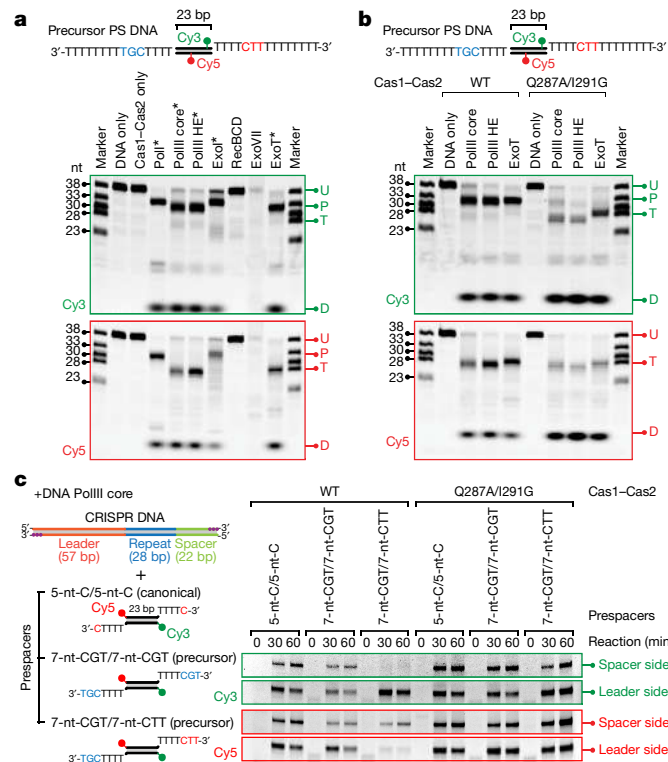


Fig. 3 | Delayed PAM trimming results in integration of spacers in the correct orientation. **a**, In vitro trimming assay using wild-type Cas1–Cas2 and 3'-5' exonuclease candidates. DnaQ or DnaQ-like enzymes are marked with an asterisk. **b**, In vitro trimming assay using wild-type (WT) Cas1–Cas2 and mutant Cas1(Q287A/I291G)–Cas2 with DNA PolIII and ExoT. Precursor prespacer DNA substrates are shown on top, and untrimmed (U), partially trimmed (P), trimmed (T) and degraded (D) products are indicated (**a**, **b**). Samples were collected after a 30-min incubation with exonucleases. **c**, In vitro trimming-driven integration assay using wild-type and mutant Cas1(Q287A/I291G)–Cas2 and DNA PolIII core. Dye-labelled canonical prespacer DNA (5-nt-C/5-nt-C) and precursor prespacer DNAs without PAM (7-nt-CGT/7-nt-CGT) or with PAM (7-nt-CGT/7-nt-CTT) were used. The linear CRISPR DNA substrate was modified with three consecutive PTO linkages at both 3'-ends for protection against non-specific degradation by DNA PolIII. The products of integration at the leader side and the spacer side are 78-nt and 113-nt long, respectively. The contrast of areas of spacer-side and leader-side integration products was adjusted for optimal visibility. For gel source data, see Supplementary Fig. 1. All data are representative of three replicates with similar results.

(DNA PolIII core); the DNA PolIII holoenzyme (DNA PolIII HE); exonuclease I (ExoI); RecBCD (ExoV); exonuclease VII (ExoVII); and exonuclease T (ExoT)³². We designed an in vitro trimming assay that enabled us to distinguish the efficiencies of the 3'-5' exonuclease enzymes on both PAM-containing and PAM-deficient strands (Fig. 3a).

The trimming assay showed that Dna PolI and ExoI exhibited a weak level of trimming (Fig. 3a); the majority of products were partially trimmed (to a size of around 32 nt) and could not be used as substrates for adaptation. Notably, DNA PolIII (core or holoenzyme) and ExoT trimmed the 3'-overhang of the PAM-deficient bottom strand to the canonical size of 28 nt (5-nt 3'-overhang) (Fig. 3a, Extended Data Fig. 5). By contrast, most PAM-containing top strands were partially trimmed to a non-canonical size (around 31 nt) (Fig. 3a, Extended Data Fig. 5a). An assay using a prespacer precursor without a PAM sequence resulted in both strands being trimmed to the canonical size (28 nt) (Extended Data Fig. 5b), whereas a prespacer precursor that contained a PAM sequence in both strands showed only partially trimmed strands (around 31 nt) (Extended Data Fig. 5b). These results demonstrate that

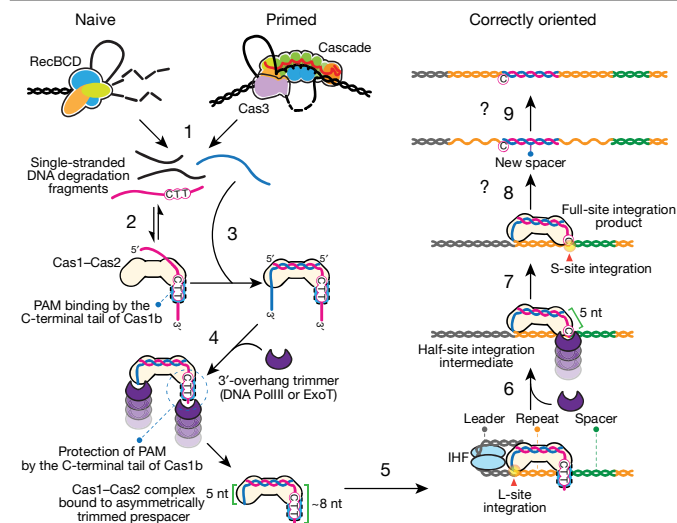


Fig. 4 | Prespacer selection, trimming and integration in naive and primed adaptation. (1) RecBCD and Cascade–Cas3 complexes generate ssDNA degradation fragments that can be used for naive and primed adaptation, respectively. (2) PAM-containing (5′-CTT-3′) ssDNA strands are captured by Cas1–Cas2. (3) Complementary strands are annealed by Cas1–Cas2. Prespacer precursors loaded into Cas1–Cas2 are likely to have 3′-overhangs longer than 5 nt. (4) A PAM-deficient strand is converted to the canonical size (5 nt) by DNA PolIII or ExoT, whereas a PAM-containing strand is partially trimmed (to a size of about 8 nt) owing to the interaction between PAM and the C-terminal tail of Cas1b. (5) The mature non-PAM-derived 3′-end is integrated at the leader side of the first repeat. (6) The partially trimmed PAM-derived 3′-overhang of the half-site intermediate is released and further trimmed into the canonical size. (7) The mature PAM-derived end is integrated at the spacer side. (8) Cas1–Cas2 disengages from the CRISPR locus by an unknown mechanism. (9) DNA repair enzymes fill the gaps, duplicating repeats. Unlike a previous model (Extended Data Fig. 9), this ‘delayed PAM trimming’ model explains the bias for correct orientation of a newly integrated spacer in the CRISPR locus.

prespacer precursors with a PAM site in only one of the two 3′-overhangs are asymmetrically trimmed by DNA PolIII and ExoT trimmers.

Biased integration by asymmetric trimming

Our data suggest that the C-terminal tail of Cas1 is essential for recognizing the PAM sequence (Fig. 1i, Extended Data Fig. 2c). We thus hypothesized that the PAM site in the 3′-overhang of prespacer precursors is protected from nuclease attacks by the interaction with the C-terminal tail of Cas1. In vitro trimming assays with the PAM-binding mutant Cas1(Q287A/I291G)–Cas2 showed that the C-terminal tail of Cas1b protects the PAM from being processed by trimmers and thereby results in prespacer DNAs with asymmetrically trimmed 3′-overhangs (Fig. 3b, Extended Data Fig. 5c, d).

From the asymmetry in prespacer maturation, we hypothesized that the canonically trimmed PAM-deficient strand (28 nt) is integrated first at the leader-side integration site (L-site), producing a half-site intermediate; this is followed by the integration of the non-canonically trimmed PAM-containing strand (31 nt) at the spacer-side integration site (S-site) of the CRISPR DNA. To test whether the asymmetry in the prespacer precursors generates a bias for correctly oriented spacers, we repeated the in vitro integration assay with asymmetrically trimmed prespacer DNAs (Extended Data Fig. 4). Whereas the symmetrically trimmed canonical prespacer DNAs were integrated at both the S-site and the L-site without showing any bias, asymmetrically trimmed prespacer DNAs showed a bias towards the integration of the 3′-end of PAM-deficient strands at the L-site (Extended Data Fig. 4). This finding supports a model of stepwise integration into the CRISPR DNA²⁰, which we validated using a smFRET assay (Extended Data Fig. 6).

Spacer orientation by delayed trimming

To complete the integration of spacers, the PAM-containing end of a prespacer precursor must be trimmed to the canonical size of 28 nt (5-nt 3′-overhang) for integration at the S-site of the CRISPR locus. To determine whether DNA PolIII core and ExoT can process PAM-containing partially trimmed 3′-overhangs, we performed an in vitro trimming-driven integration assay with half-site intermediates (Extended Data Fig. 7a). The PAM-containing top prespacer strands with various lengths of 3′-overhang were processed to the canonical size and integrated, resulting in S-site integration products (Extended Data Fig. 7b–d).

To investigate whether Cas1–Cas2 and the integration host factor (IHF) dimer, together with prespacer trimmers, are sufficient for spacers to be integrated in the correct orientation, we performed an in vitro trimming-driven integration assay with a phosphorothioate (PTO)-modified linear CRISPR DNA molecule and various prespacer precursors (Fig. 3c, Extended Data Fig. 8a). Whereas the symmetrically trimmed prespacer DNA (5-nt-C/5-nt-C) resulted in a comparable ratio of correctly and incorrectly oriented full-site integration products, the asymmetrically trimmed (5-nt-C/15-nt-CTT), partially trimmed (7-nt-CGT/7-nt-CTT) and untrimmed (15-nt-CGT/15-nt-CTT) prespacer precursors resulted in full-site integration products with a bias for the PAM-containing end towards the S-site of the CRISPR DNA (Fig. 3c, Extended Data Fig. 8b). We further confirmed the trimming-driven integration using a smFRET assay (Extended Data Fig. 8c, d). To clarify the role of PAM protection by the Cas1 C-terminal tail in this orientation bias, we repeated the in vitro trimming-driven integration assay using the PAM-binding mutant Cas1(Q287A/I291G)–Cas2, which exhibited no bias in the integration orientation (Fig. 3c). Therefore, delayed PAM trimming of precursor prespacer DNAs results in a strong bias for the integration of correctly oriented spacers in the CRISPR locus, which confers robustness to CRISPR–Cas immunity (Fig. 4).

Discussion

Until now, the working model for the acquisition of spacers has described prespacer binding, maturation and integration as independent steps (Extended Data Fig. 9). This model is incomplete, however, as it necessitates symmetrically trimmed canonical prespacer DNA for integration—which cannot explain how the cytosine residue at the 3′-end of the PAM sequence is correctly oriented for spacer integration. We have shown here that the processes of prespacer binding, maturation and integration are tightly coordinated in time (Fig. 4). Cas1–Cas2 makes frequent interactions with suitable prespacer precursors among a wide variety of substrates, including ssDNA and partially duplexed DNA with long 3′-overhangs. Once loaded, DNA PolIII and other DnaQ-like exonucleases process prespacer precursors into the mature versions, while the PAM is protected by the C-terminal tail of Cas1. This results in an asymmetric maturation of the prespacer precursors and a bias for L-site integration of the non-PAM end of the prespacer. Next, the PAM end of the half-site-intermediate prespacer is released from Cas1, processed to the mature prespacer and subsequently integrated into the S-site, resulting in correctly oriented and functional spacers in the CRISPR loci. This asymmetry in trimming and subsequent biased integration may be equally important in dictating correct positioning in other types of CRISPR–Cas systems (type II, for example), but different proteins could be involved in PAM recognition and precursor trimming.

Our single-molecule assay with ssDNA substrates revealed that Cas1–Cas2 can capture PAM-containing ssDNA fragments and facilitate the formation of prespacer precursors by recruiting the complementary strand. On the basis of these findings, we propose a model in which Cas1–Cas2 captures ssDNA fragments in the cell to acquire functional prespacer substrates. In this model, Cas1–Cas2 transiently interacts with ssDNA through a facilitated diffusion mechanism to search for a PAM sequence. Cas1–Cas2 has been shown to form a complex with

Cascade–Cas3 during primed spacer acquisition^{33,34}. The formation of this complex might allow ssDNA fragments to be directly transferred to Cas1–Cas2, ensuring that primed spacer acquisition is robust. Single-molecule studies have shown that Cas3 generates ssDNA loops during CRISPR interference^{33,35}. It will be interesting to ascertain whether ssDNA loops that are marked with PAM sequences can be directly recognized by the Cas1–Cas2 complex (Extended Data Fig. 10). In addition, another study suggested that Chi sites are hotspots for the naive acquisition of new spacer candidates *in vivo*¹². In the areas around Chi sites, proximal RecBCD forms long ssDNA loops, which might also act as initial docking regions for Cas1–Cas2 (Extended Data Fig. 10).

The fact that DNA PolIII has DNA polymerization activity suggests that there might be a molecular link between the trimming of prespacer precursors and the duplication of repeats, through the trimming-driven integration of prespacers. However, the intracellular copy numbers of the DNA PolIII holoenzyme are low (10–20 copies per cell) and its expression level is tightly regulated³⁶. This necessitates the involvement of abundant 3′–5′ exonucleases for efficient processing of prespacer precursors, and hence allows the CRISPR–Cas immune system to keep pace with rapid infections. The contribution of ExoT may compensate for the limited availability of DNA PolIII during the maturation of prespacer precursors, and it is likely that several other unidentified exonucleases are also involved.

Several groups have harnessed the nucleic-acid-acquisition abilities of Cas1–Cas2 to develop new techniques for recording nucleic acids in cellular contexts³⁷. Cas1–Cas2-based recording techniques allow cellular events in prokaryotes to be captured in chronological order³⁷. Our results may help in developing a next generation of Cas1–Cas2 recorders that are more efficient at capturing information. Moreover, our findings may also enable a Cas1–Cas2-based recording system to be developed in eukaryotes, which has not previously been reported.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2018-1>.

- Hille, F. et al. The biology of CRISPR–Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
- Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* **527**, 535–538 (2015).
- Nuñez, J. K. et al. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).
- Nuñez, J. K., Lee, A. S., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **519**, 193–198 (2015).
- Wang, J. et al. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR–Cas Systems. *Cell* **163**, 840–853 (2015).
- Jackson, S. A. et al. CRISPR–Cas: adapting to change. *Science* **356**, eaal5056 (2017).
- McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol.* **17**, 7–12 (2019).
- Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
- Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).

- Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
- Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
- Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015).
- Künne, T. et al. Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell* **63**, 852–864 (2016).
- Musharova, O. et al. Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* **45**, 3297–3307 (2017).
- Semenova, E. et al. Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR–Cas interfering complex. *Proc. Natl Acad. Sci. USA* **113**, 7626–7631 (2016).
- Yeeles, J. T., Gwynn, E. J., Webb, M. R. & Dillingham, M. S. The AddAB helicase-nuclease catalyses rapid and processive DNA unwinding using a single superfamily 1A motor domain. *Nucleic Acids Res.* **39**, 2271–2285 (2011).
- Yeeles, J. T., van Aelst, K., Dillingham, M. S. & Moreno-Herrero, F. Recombination hotspots and single-stranded DNA binding proteins couple DNA translocation to DNA unwinding by the AddAB helicase-nuclease. *Mol. Cell* **42**, 806–816 (2011).
- Mulepati, S. & Bailey, S. In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* **288**, 22184–22192 (2013).
- Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR immunological memory requires a host factor for specificity. *Mol. Cell* **62**, 824–833 (2016).
- Xiao, Y., Ng, S., Nam, K. H. & Ke, A. How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature* **550**, 137–141 (2017).
- Lopez-Sanchez, M. J. et al. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol. Microbiol.* **85**, 1057–1071 (2012).
- Shmakov, S. et al. Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.* **42**, 5907–5916 (2014).
- Shimori, M., Garrett, S. C., Graveley, B. R. & Terns, M. P. Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Mol. Cell* **70**, 814–824 (2018).
- Rolfe, C., Graham, S., Rouillon, C. & White, M. F. Prespacer processing and specific integration in a type I-A CRISPR system. *Nucleic Acids Res.* **46**, 1007–1020 (2018).
- Lee, H., Zhou, Y., Taylor, D. W. & Sashital, D. G. Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. *Mol. Cell* **70**, 48–59 (2018).
- Kieper, S. N. et al. Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.* **22**, 3377–3384 (2018).
- Hou, Z. & Zhang, Y. Insights into a mysterious CRISPR adaptation factor, Cas4. *Mol. Cell* **70**, 757–758 (2018).
- Moch, C., Fromant, M., Blanquet, S. & Plateau, P. DNA binding specificities of *Escherichia coli* Cas1–Cas2 integrase drive its recruitment at the CRISPR locus. *Nucleic Acids Res.* **45**, 2714–2723 (2017).
- Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. & Severinov, K. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol.* **10**, 716–725 (2013).
- Shiryaeva, A. A. et al. Detection of spacer precursors formed *in vivo* during primed CRISPR adaptation. *Nat. Commun.* **10**, 4603 (2019).
- Drabavicius, G. et al. DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR–Cas system. *EMBO Rep.* **19**, e45543 (2018).
- Lovett, S. T. The DNA exonucleases of *Escherichia coli*. *Ecosal Plus* **4**, <https://doi.org/10.1128/ecosalplus.4.4.7> (2011).
- Dillard, K. E. et al. Assembly and translocation of a CRISPR–Cas primed acquisition complex. *Cell* **175**, 934–946 (2018).
- Redding, S. et al. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR–Cas system. *Cell* **163**, 854–865 (2015).
- Loeff, L., Brouns, S. J. J. & Joo, C. Repetitive DNA reeling by the Cascade–Cas3 complex in nucleotide unwinding steps. *Mol. Cell* **70**, 385–394 (2018).
- Wu, Y. H., Franden, M. A., Hawker, J. R. Jr & McHenry, C. S. Monoclonal antibodies specific for the alpha subunit of the *Escherichia coli* DNA polymerase III holoenzyme. *J. Biol. Chem.* **259**, 12117–12122 (1984).
- Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Protein preparation

Cas1–Cas2 complex was expressed in *E. coli* BL21-AI chemically competent cells (Thermo Fisher Scientific, C607003) using pET52b cloned with the Cas1–Cas2 operon with Strep-tag II (N-terminal) (pWUR871) and purified as described¹³. In brief, cells were grown to an optical density at 600 nm (OD_{600}) of 0.4 in Luria-Bertani (LB) broth medium, cooled on ice for 30 min and induced with 0.5 mM IPTG (Gold Biotechnology, I2481C50) and 0.2% L-arabinose (Gold Biotechnology, A-300-1). Protein expression was induced overnight at 20 °C. Cells were collected by centrifugation and lysed in 20 mM HEPES–NaOH pH 7.5 (HEPES, Sigma, H3375), 75 mM NaCl (Sigma, S9888), 1 mM DTT (Gold Biotechnology, DTT100), 5% glycerol (Sigma, G7893-1L) and 0.1% Triton X-100 (Sigma, X100-500ML) using a Stansted pressure-cell homogenizer. The lysate was cleared by centrifugation and incubated with strep-tactin sepharose (Iba Lifesciences, 2-1201-025) for 1 h at 4 °C. Next, the lysate with beads was loaded onto a gravity column and washed with 20 mM HEPES–NaOH pH 7.5, 300 mM NaCl, 1 mM DTT and 5% glycerol, followed by elution with a buffer that contained 20 mM HEPES–NaOH pH 7.5, 75 mM NaCl, 1 mM DTT and 5% glycerol (Sigma-Aldrich, G7893-1L) (storage buffer) with 4 mM D-desthiobiotin (Iba Lifesciences, 2-1000-005). The presence and purity of the Cas1–Cas2 complex was checked through Bis-Tris 4–12% NuPAGE (Thermo Fisher Scientific, NP032A) with NuPAGE MES SDS running buffer (Thermo Fisher Scientific, NP0002). The protein complex was concentrated with Amicon ultra centrifugal filters (Merck Millipore) and further purified on a Superdex 200 10/300 GL size-exclusion column (GE Healthcare, 17517501) using the ÄKTA pure protein purification system (GE Healthcare). The final complex was diluted in storage buffer with 50% glycerol, snap-frozen in liquid nitrogen and stored at –80 °C. Mutant Cas1–Cas2 complexes were cloned by site-directed mutagenesis as described previously³⁸ and purified as described above. Primers for sub-cloning are listed in Supplementary Table 1. The *ihfA* and *ihfB* genes were PCR amplified from *E. coli* (BL21) genomic DNA using the indicated primers in Supplementary Table 1, and subsequently cloned into Berkeley MacroLab ligation-independent cloning (LIC) vectors 13K-HR (Addgene, plasmid 48318) and 13S-A (Addgene, plasmid 48323), respectively, as described previously^{39,40}. The *ihfA* gene was cloned to code for an N-terminal His₆-tagged IHF α protein upon expression. Subsequent steps for the purification of the IHF α and IHF β dimer complex were performed using HisPur Ni-NTA Resin (Thermo Fisher Scientific, 88222) as previously described¹⁹. DNA PolIII proteins were purified as described previously^{41–43}.

For site-specific Cas1–Cas2 biotinylation, an N-terminal LCTPSR formylglycine-generating enzyme (FGE) recognition motif on Cas1 was inserted by site-directed mutagenesis in plasmid pWUR871 with the primers shown in Supplementary Table 1, and co-expressed with FGE proteins (Addgene, plasmid 16132)⁴⁴. The purified Cas1–Cas2 complex was buffer-changed with 0.5 M sodium acetate (pH 5.5), labelled with EZ-Link Biotin-LC-Hydrazide (Thermo Fisher Scientific, 21340) and incubated overnight at room temperature. Labelled Cas1–Cas2 was purified by size-exclusion chromatography with Superdex 200 10/300 GL. Fractions were concentrated using Amicon Ultra-4 centrifugal filters MWCO 30K (Merck, UFC803024), pooled in storage buffer with 50% glycerol, snap-frozen in liquid nitrogen and stored at –80 °C.

DNA preparation

Synthetic DNA oligonucleotides (Ella Biotech) were internally labelled with a monoreactive *N*-hydroxysuccinimide (NHS)-ester form of cyanine dyes as donors (Cy3 mono-reactive NHS ester, GE Healthcare,

PA13101) or acceptors (Cy5 mono-reactive NHS ester, GE Healthcare, PA15101), or EZ-Link NHS-biotin (Thermo Fisher Scientific, 20217) at amino-C6-dT (amine modification with amino-modifier C6-T) (Supplementary Table 1). After labelling, the ssDNA strands were annealed in 20 mM Tris (pH 8.0), 150 mM KCl and 5 mM MgCl₂ using a thermocycler (Bio-Rad) at –1 °C per 1-min cycle from 95 °C to 16 °C, and then stored at 4 °C.

Single-molecule TIRF imaging and data acquisition

The fluorescent labels Cy3 and Cy5 were imaged using prism-type total internal reflection microscopy³⁵. In brief, Cy3 was imaged through excitation by a 532-nm diode laser (Compass 215M-50, Coherent). Cy5 was detected by FRET with Cy3, but, if necessary, Cy5 was directly excited using a 640-nm solid-state laser (CUBE 640-100C, Coherent). Fluorescence signals from single molecules were collected through a 60 \times water-immersion objective (UPlanSApo, Olympus) with an inverted microscope (IX71, Olympus). Scattering of the 532-nm laser beam was blocked with a 550-nm long-pass filter (LP03-532RU-25, SemRock). When the 640-nm laser was used, 640-nm laser scattering was blocked with a notch filter (633 ± 12.5 nm, NF03-633E-25, SemRock). Subsequently, signals of Cy3 and Cy5 were spectrally split with a dichroic mirror ($\lambda^{\text{cut-off}} = 635$ nm, Chroma) and imaged onto halves of an electron-multiplying CCD camera (iXon 897, Andor Technology).

To eliminate non-specific surface adsorption of proteins and nucleic acids to a quartz surface (Finkenbeiner), piranha-etched slides were PEG-passivated over two rounds of PEGylation⁴⁵. After assembly of a microfluidic flow chamber, slides were incubated for 10 min with T50 buffer (Tris–HCl pH 8.0 (Sigma-Aldrich, Trizma base, T6066) and 50 mM NaCl) containing 5% Tween-20 (Sigma-Aldrich, P7949) to further improve slide quality⁴⁶. Next, the chamber was incubated with 20 μ l of 0.1 mg ml^{–1} streptavidin (Invitrogen, S-888) for 5 min followed by a washing step with 100 μ l Cas1–Cas2 buffer (50 mM HEPES–NaOH pH 7.5, 50 mM KCl (Ambion, AM9530G) and 5 mM MgCl₂). Biotinylated Cas1–Cas2 at 0.2–1 nM was specifically immobilized through biotin–streptavidin linkage by incubating the chamber for 5 min. Remaining unbound biotin–Cas1–Cas2 was flushed away with 100 μ l Cas1–Cas2 imaging buffer (50 mM HEPES–NaOH pH 7.5, 50 mM KCl, 5 mM MgCl₂ (Ambion, AM9640G), glucose oxidase (Sigma, G2133), 4 mg ml^{–1} catalase (Roche, 10106810001) and 1 mM Trolox ((+/-)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid, Sigma, 238813). Immobilized Cas1–Cas2 was incubated with 0.5–2 nM labelled DNA at room temperature (23 ± 1 °C) for the indicated times.

To visualize the dynamics of prespacer and prespacer precursor DNA binding on Cas1–Cas2, Cy3 molecules were excited on an area of 50×50 μ m² with a green laser (532 nm) at 28% of the full laser power (9 mW), and the time resolution was set to 0.1 s. Under these imaging conditions we obtained a high signal-to-noise ratio that allowed us to visualize kinetic intermediates while imaging over time periods of 3.5 min. Under these conditions, photobleaching of the donor and acceptor dyes during our observation time was minimized.

Single-molecule data analysis

A series of CCD images were acquired with laboratory-made software at a time resolution of 0.1 s. Fluorescence time traces were extracted with an algorithm written in IDL (ITT Visual Information Solutions) that picked fluorescence spots above a threshold with a defined Gaussian profile. The extracted time traces were analysed using custom written MATLAB (MathWorks) algorithms. FRET efficiency was defined as the ratio between the acceptor intensity and the sum of the acceptor and donor intensities. To determine the dissociation rate (k_{off}), the start and end of each binding event were determined (Fig. 1b). The start of each event was marked by an abrupt increase in the fluorescence signal, whereas the end of each event was marked by an abrupt decrease (Fig. 1b). Selecting the start and end of each event yielded the duration of each event, which was plotted in a histogram.

These dwell-time distributions were fitted with a single-exponential decay using maximum-likelihood estimations (Extended Data Figs. 1c, i, m, 3c). This fit yielded the average dwell-time (τ_{off}), which was then converted to the dissociation frequency (rate) ($k_{\text{off}} = 1/\tau_{\text{off}}$). The 95% CIs (errors) of the dissociation frequencies were obtained by empirical bootstrap analysis⁴⁷.

The binding frequency was determined by measuring the time from flow-in of the DNA substrate to the occurrence of the first binding event. These characteristic times were plotted as a cumulative histogram and fitted with a single-exponential decay using maximum-likelihood estimation (Extended Data Figs. 1h, j, l, 2d, 3b). This fit yielded the average arrival time (τ_{on}), which was then converted to the binding frequency ($k_{\text{on}} = 1/(\tau_{\text{on}}c)$), where c is the concentration of DNA. The 95% confidence intervals (errors) of the binding frequencies were obtained by empirical bootstrap analysis⁴⁷.

To obtain survival rates of the long-lived population, fluorescently labelled prespacer DNA was incubated for 10 min in the microfluidic chamber. After washing the remaining unbound molecules ($t = 0$), the bound population was tracked over a time of 45 min. To avoid photobleaching, short snapshots of 10 frames were taken over 20 fields of view at each time point, providing the average number of molecules bound to Cas1–Cas2. For the subsequent analysis, the number of lost molecules at each time point was subtracted from the total number of molecules bound at $t = 0$. This yielded survival rate curves that were fitted with a single-exponential decay (Extended Data Fig. 1n).

Quantification and statistical analysis

Histograms and fits were generated using OriginPro (OriginLab). The averages and errors representing the number of bound molecules (Fig. 1f, Extended Data Fig. 2b), the survival time (Fig. 1g, Extended Data Fig. 1n), the cumulative probability of the molecule arrival time (Figs. 1d, h, i, 2a, d, Extended Data Figs. 1h, j, l, 2d, 3b, d), the average dwell time of events (Figs. 1e, h, i, 2b, Extended Data Figs. 1c, i, m, 3c, e) and the FRET population analysis (Extended Data Figs. 1b, f, g, p, q, 2f, g, 6e, 8d) encompass a minimum of three replicates (n). The errors represent s.e.m., which was defined as s.e.m. = σ/\sqrt{n} . The averages and errors displayed in the figures were obtained through bootstrap analysis. In brief, for bootstrap analysis, 10^4 datasets were generated by random sampling with replacement from the original dataset. Each of these datasets was fitted with the respective fit (indicated in the figure legend) and then used to calculate the average and 95% CI, which was defined as $\text{CI}(95) = 1.96\sigma$.

In vitro integration assay

Integration assays with fluorescently labelled prespacer and prespacer precursor DNA were performed using 200 nM Cas1–Cas2 complex, 500 nM IHF dimer, 20 nM prespacer or prespacer precursor DNA and 40 nM CRISPR DNA in a final reaction with Cas1–Cas2 integration buffer (50 mM HEPES–NaOH pH 7.5, 50 mM KCl, 5 mM MgCl₂ and 5% PEG8000 (Sigma-Aldrich, P2139)). The Cas1–Cas2 complex was incubated with prespacer or prespacer precursor DNA to allow complex formation for 30 min at room temperature. Subsequently, IHF dimer was incubated with CRISPR DNA in a separate tube. The reaction was activated by adding the prespacer–Cas1–Cas2 complex to the IHF–CRISPR DNA mix, and incubated for the indicated times at 37 °C. For full-site integration experiments of half-site intermediate constructs, 20 nM half-site integration products were incubated with 500 nM IHF dimer for 30 min at room temperature, followed by the addition of 200 nM Cas1–Cas2. This mixture was incubated at room temperature for 10 min, allowing Cas1–Cas2 and IHF dimer to assemble with the DNA. Next, 100 nM DNA PolIII core, 2.5 U ExoT (New England Biolabs, M0265) or no enzyme was added, followed by incubation at 37 °C for the indicated times. For trimming-driven integration assays, prespacer–Cas1–Cas2 and IHF–CRISPR DNA mixtures were combined and incubated at room temperature for 10 min. The combined mixtures were added with

100 nM DNA PolIII core or 2.5 U ExoT, followed by incubation at 37 °C for the indicated times. To quench the reactions, DNA loading buffer (final concentration 12.5 mM EDTA (Ambion, AM9261) and 47.5% formamide (Roche, 11814320001)) was added and thoroughly mixed with samples. The samples were heated at 95 °C for 10 min and immediately loaded and run on 15 × 15-cm²-sized 7 M urea (Sigma-Aldrich, 316830) denaturing 9% or 12% polyacrylamide 1× Tris-borate-EDTA (TBE) gels. The gels were pre-run for 2 h and run for 2–3 h at 370 V in 0.5× TBE buffer (Promega, V4251). Fluorescence signals from gels were analysed in an Amersham Typhoon biomolecular imager.

In vitro trimming assay

For in vitro trimming assays, 20 nM prespacer or prespacer precursor DNA was incubated with 200 nM Cas1–Cas2 in Cas1–Cas2 integration buffer containing 10 mM DTT and 10% PEG8000 at room temperature for 30 min, and then supplied with each exonuclease in the indicated amounts: 1 U DNA PolI (New England Biolabs, M0209), 100 nM DNA PolIII core (for holoenzyme, 33.3 nM clamp loader, 200 nM β -clamp and 10 nM DnaBC helicase were also added, but without single-stranded DNA binding protein (SSB)), 1 U ExoI (New England Biolabs, M0293), 1 U RecBCD (New England Biolabs, M0345), 0.5 U ExoVII (New England Biolabs, M0379) or 1 U ExoT. After incubation at 37 °C for the indicated times, the reaction was quenched with DNA-loading buffer and subsequently heated at 95 °C for 10 min for 15 × 15-cm²-sized 7 M urea denaturing 20% TBE–PAGE. Fluorescence signals from gels were analysed in an Amersham Typhoon biomolecular imager.

Electromobility shift assay

Binding assays were performed in buffer containing 50 mM HEPES–NaOH pH 7.5, 50 mM KCl, 5 mM MgCl₂, 5% PEG8000, 5% glycerol and 1 mM DTT without (Extended Data Fig. 1r) or with (Extended Data Fig. 3g) 25 mM EDTA. Each reaction contained 10 nM dye-labelled prespacer DNAs or ssDNAs at increasing concentrations (0–200 nM) of Cas1–Cas2. The reactions were incubated at room temperature for 30 min and resolved at 4 °C on 4% native agarose (Promega, V3121) gels containing 1× Tris-acetate-EDTA (TAE) buffer (Promega, H5231). Fluorescence signals from gels were analysed in an Amersham Typhoon biomolecular imager.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data generated or analysed during this study are either included in this Article and its Supplementary Information or available from the corresponding authors on request. Source Data for Figs. 1, 2 and Extended Data Figs. 1–3, 6, 8 are provided with the paper.

38. Kim, S. et al. Temporal landscape of microRNA-mediated host–virus crosstalk during productive human cytomegalovirus infection. *Cell Host Microbe* **17**, 838–851 (2015).
39. Thieme, F., Engler, C., Kandzia, R. & Marillonnet, S. Quick and clean cloning: a ligation-independent cloning strategy for selective cloning of specific PCR products from non-specific mixes. *PLoS One* **6**, e20556 (2011).
40. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074 (1990).
41. Jergic, S. et al. A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. *EMBO J.* **32**, 1322–1333 (2013).
42. Hamdan, S. et al. Hydrolysis of the 5'-p-nitrophenyl ester of TMP by the proofreading exonuclease (e) subunit of *Escherichia coli* DNA polymerase III. *Biochemistry* **41**, 5266–5275 (2002).
43. Lewis, J. S. et al. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *eLife* **6**, e23932 (2017).
44. Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321–322 (2007).
45. Chandradoss, S. D. et al. Surface passivation for single-molecule protein studies. *J. Vis. Exp.* **86**, e50549 (2014).

Article

46. Pan, H., Xia, Y., Qin, M., Cao, Y. & Wang, W. A simple procedure to improve the surface passivation for single molecule fluorescence studies. *Phys. Biol.* **12**, 045006 (2015).
47. Dekking, M. *A Modern Introduction to Probability and Statistics: Understanding Why and How* (Springer, 2005).

Acknowledgements We thank A. C. Haagsma and T. Künne for providing Cas1–Cas2 vectors and proteins; S. Leachman, N. Dekker and the members of the C.J. and S.J.J.B. laboratories for discussions; and T. J. Cui for discussions on kinetic models. S.J. thanks N. Dixon for guidance. S.K. was partly funded by a Marie Skłodowska-Curie grant (753528); C.J. and S.J.J.B. were funded by the Foundation for Fundamental Research on Matter (15PR3188); and S.J. was funded by a collaborative grant from King Abdullah University of Science and Technology, Saudi Arabia (OSR-2015-CRG4-2644).

Author contributions S.K., L.L., S.J.J.B. and C.J. conceived the study. S.K. and S.C. purified and labelled Cas1–Cas2 proteins. S.K. and L.L. performed biochemical experiments. S.J. purified DNA PolIII and its related proteins. S.K. and L.L. performed smFRET experiments and analysed data. S.K., L.L., S.J.J.B. and C.J. discussed the data and wrote the manuscript.

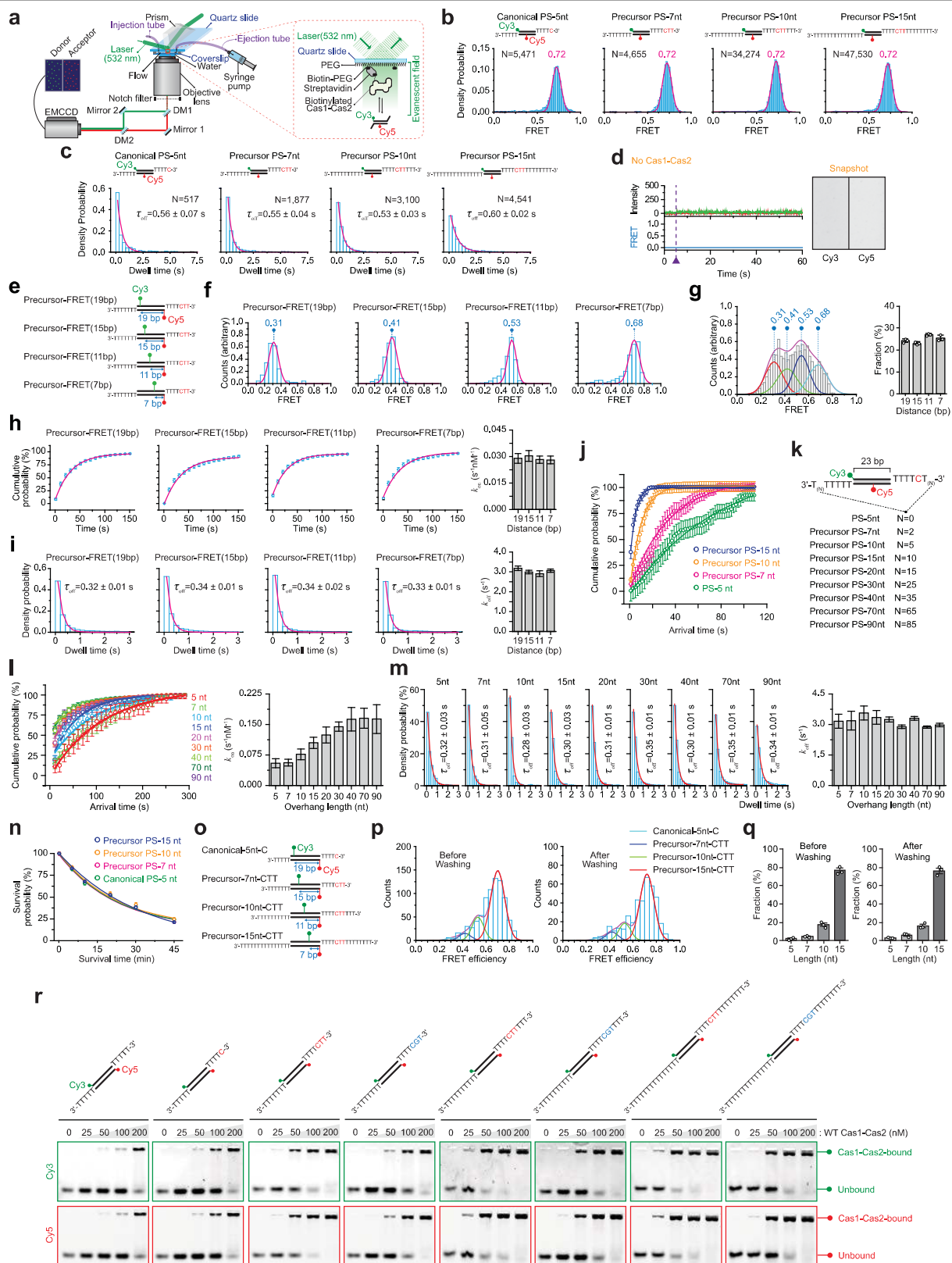
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2018-1>.

Correspondence and requests for materials should be addressed to S.K. or C.J.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



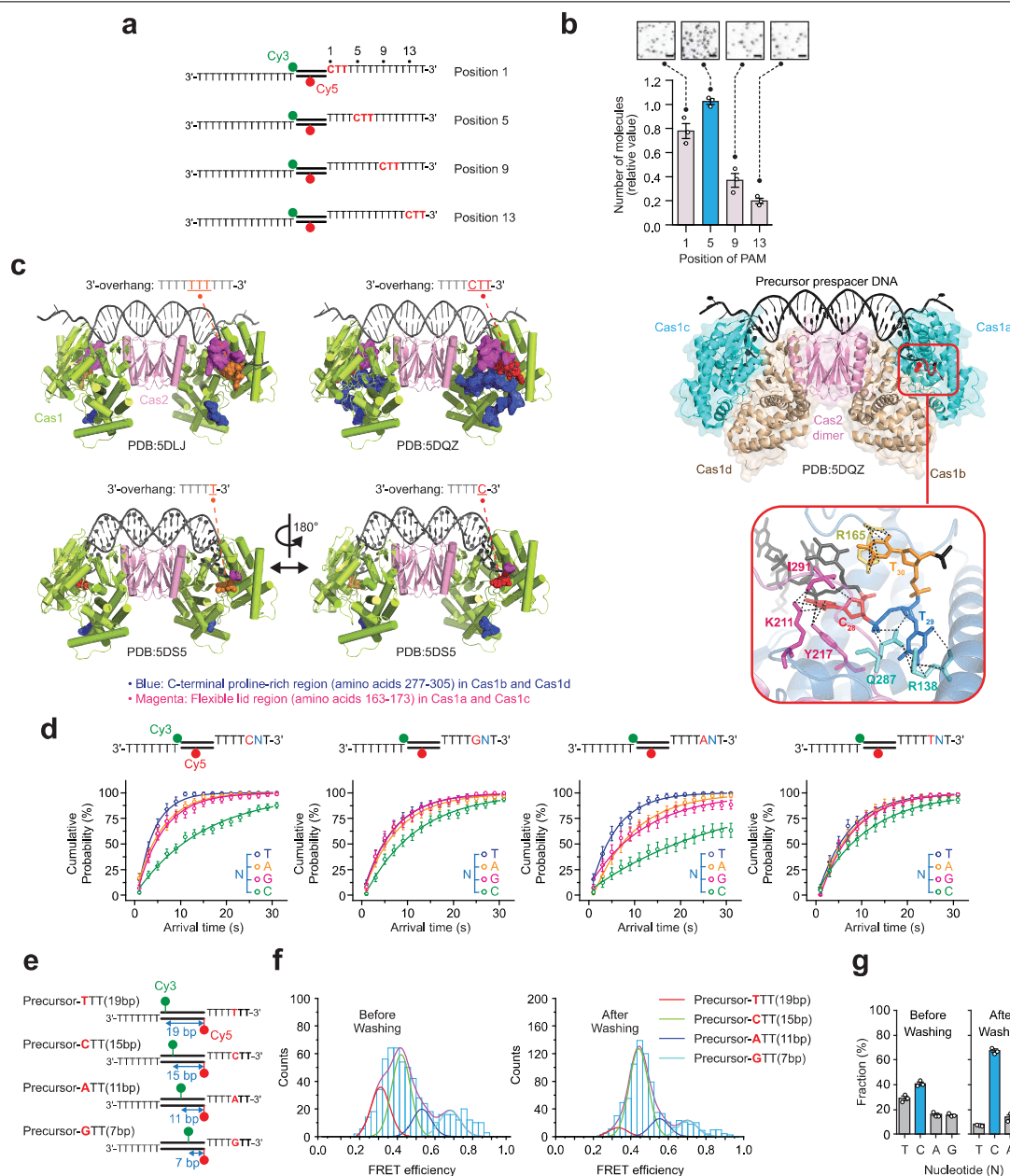
Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | Single-molecule and biochemical analysis of the effect of 3'-overhang length.

a, Schematic of the single-molecule TIRF set-up that was used for measuring Cas1–Cas2 binding to canonical or precursor prespacer DNA. **b**, FRET efficiency histograms of binding events observed for canonical and precursor prespacer DNAs with 3'-overhangs of various lengths. **c**, Dwell-time (Δt) distributions and average binding time (τ_{off}) determination for canonical and precursor prespacer DNAs with 3'-overhangs of various lengths. **d**, Left, representative time trace from a binding assay in the absence of Cas1–Cas2. DNA was added at $t = 5$ s. Right, image of a field of view with Cy3 signals on the left and Cy5 signals on the right. The image was recorded 1 min after the addition of DNA. **e**, Schematics of precursor prespacer DNAs with different labelling positions. **f**, FRET efficiency histograms of individual precursor prespacer DNAs with different labelling positions bound to Cas1–Cas2. **g**, FRET distribution and fractions of precursor prespacer DNAs from a single-molecule competition assay using precursor prespacer DNAs with different labelling positions. Histograms were obtained by incubating equal concentrations of precursor prespacer DNAs. To track the stably bound population, the flow chamber was washed and the fluorescence signals of the remaining population were measured. **h**, Cumulative distribution of the arrival times for binding events and k_{on} for precursor prespacer DNAs with different labelling positions. **i**, Dwell-time distributions and k_{off} of binding events for precursor prespacer DNAs with distinct labelling positions. **j**, Cumulative probability of the arrival times for precursor prespacer DNAs with 3'-overhangs of various lengths. **k**, Schematics of precursor prespacer DNAs with 3'-overhangs of various lengths. Each DNA construct consists of a 23-bp central duplex and 5-nt 3'-overhangs at both ends, which are further extended with N number of single-stranded deoxythymidine (dT) nucleotides. Both strands were labelled with a Cy3 fluorophore at the 5'-end of the top strand and a Cy5

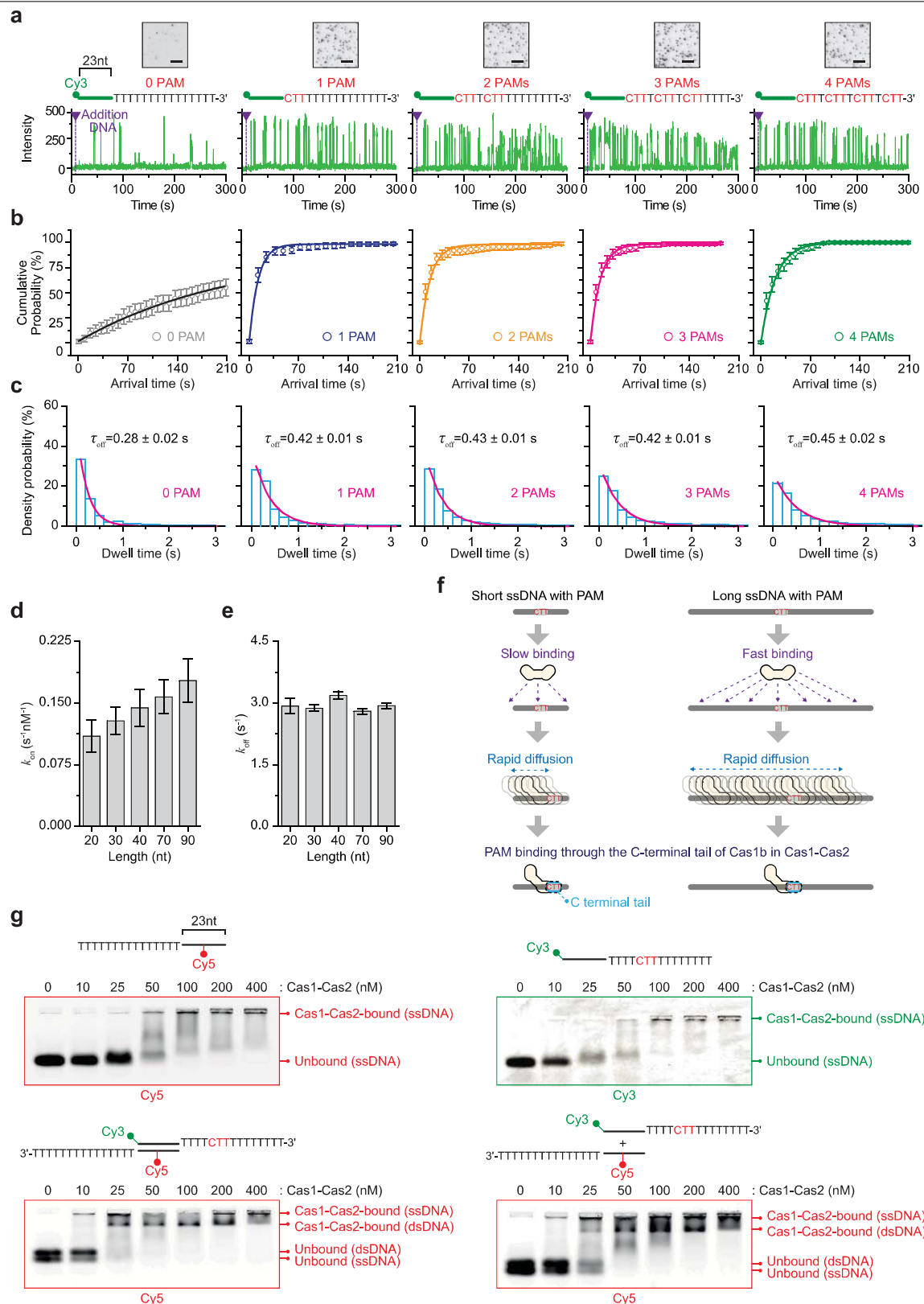
fluorophore at the 16th nucleotide (T) from the 5'-end. **l**, Cumulative distribution of the arrival times for binding events and k_{on} for precursor prespacer substrates with 3'-overhangs of various lengths. **m**, Dwell-time distributions of binding events and k_{off} for precursor prespacer DNAs with 3'-overhangs of various lengths. **n**, Survival probability of stably bound substrates with 3'-overhangs of various lengths. The solid lines represent single-exponential fits using maximum-likelihood estimation. **o**, Schematics of canonical and precursor prespacer DNAs with different labelling positions and 3'-overhang lengths for a single-molecule competition experiment. **p**, **q**, FRET distributions (**p**) and fractions (**q**) of each FRET population from single-molecule competition experiments for canonical and precursor prespacer DNAs with 3'-overhangs of various lengths. 'Before washing' includes both transient and stably bound molecules; 'after washing' includes only the stably bound molecules. **r**, EMSAs on various canonical and precursor prespacer DNA substrates with increasing amounts of wild-type Cas1–Cas2. The top and bottom strands were labelled at the 5'-end with Cy3 and Cy5, respectively. Cas1–Cas2-bound and unbound precursor prespacer DNAs are indicated on the right. For **b**, **f**, **g**, **p**, solid lines represent Gaussian fits; the centre of each peak corresponds to the predetermined position of each individual construct in **g**, **p**. For **c**, **h–j**, **l**, **m**, solid lines represent single-exponential fits (maximum-likelihood estimation) that were used to determine the binding frequency (k_{on}) (**h**, **j**, **l**) and dissociation rate (k_{off}) (**c**, **i**, **m**). For the cumulative probability of the arrival times for k_{on} bar plots and dwell times for k_{off} bar plots, data are mean \pm 95% CI, obtained by bootstrap analysis of a single replicate with $n \geq 300$ (**h**, **j**, **l**) or $n \geq 500$ (**c**, **i**, **m**) individual molecules. For the FRET fractions, data are mean \pm s.e.m. from three independent measurements ($n = 3$) with $n \geq 5,000$ molecules for each measurement (**g**, **q**). Data are representative of three replicates with similar results (**b**, **c**, **f**, **h–j**, **l–n**, **q**).



Extended Data Fig. 2 | Single-molecule analysis of the effect of the PAM position and sequence. **a**, Schematics of precursor prespacer DNA with the PAM sequence at different positions in the 3'-overhang. **b**, Average number of molecules bound per field of view after a 30-min incubation with precursor prespacer DNAs. Data are mean \pm s.e.m. ($n = 3$). Representative CCD images (acceptor channel) are included as insets. Scale bars, 5 μ m. **c**, Structural comparison of Cas1-Cas2 precursor prespacer complexes. Cas1-Cas2 in complex with a non-PAM (5'-TTT-3', orange)-containing substrate with 10-nt 3'-overhangs (PDB: 5DLJ); a PAM (5'-CTT-3', red)-containing substrate with 8-nt 3'-overhangs (PDB: 5DQZ); and a non-PAM substrate with 5-nt 3'-overhangs that end with a T (PDB: 5DS5; left) or a C (PDB: 5DS5; right). The C-terminal proline-rich tail of Cas1b and the flexible internal lid-like loop region of Cas1a are highlighted in blue and magenta, respectively. The magnified image on the right represents the molecular architecture of the PAM-recognizing residues of Cas1-Cas2. The residues of the PAM sequence (C₂₈, red; T₂₉, blue; T₃₀, orange)

are coloured, together with the PAM-interacting residues of Cas1.

d, Cumulative probability of the arrival times for precursor prespacer DNAs with different PAM sequences. A single-exponential fit (solid line) was used to determine the binding frequency (k_{on}). Data are mean \pm 95% CI, obtained by bootstrap analysis from a single replicate with $n \geq 300$ individual molecules. **e**, Schematics of the design of precursor prespacer DNAs with different PAM sequences. **f, g**, FRET distributions (**f**) and fractions (**g**) of each FRET population from single-molecule competition experiments for different sequences at the PAM position. Each population was fitted with a Gaussian distribution (solid line); the centre of each peak corresponds to the predetermined position of each individual construct (**f**). For the fractions, data are mean \pm s.e.m. from three independent measurements ($n = 3$) with $n \geq 5,000$ molecules per each measurement (**g**). Data are representative of three replicates with similar results (**d, f, g**).



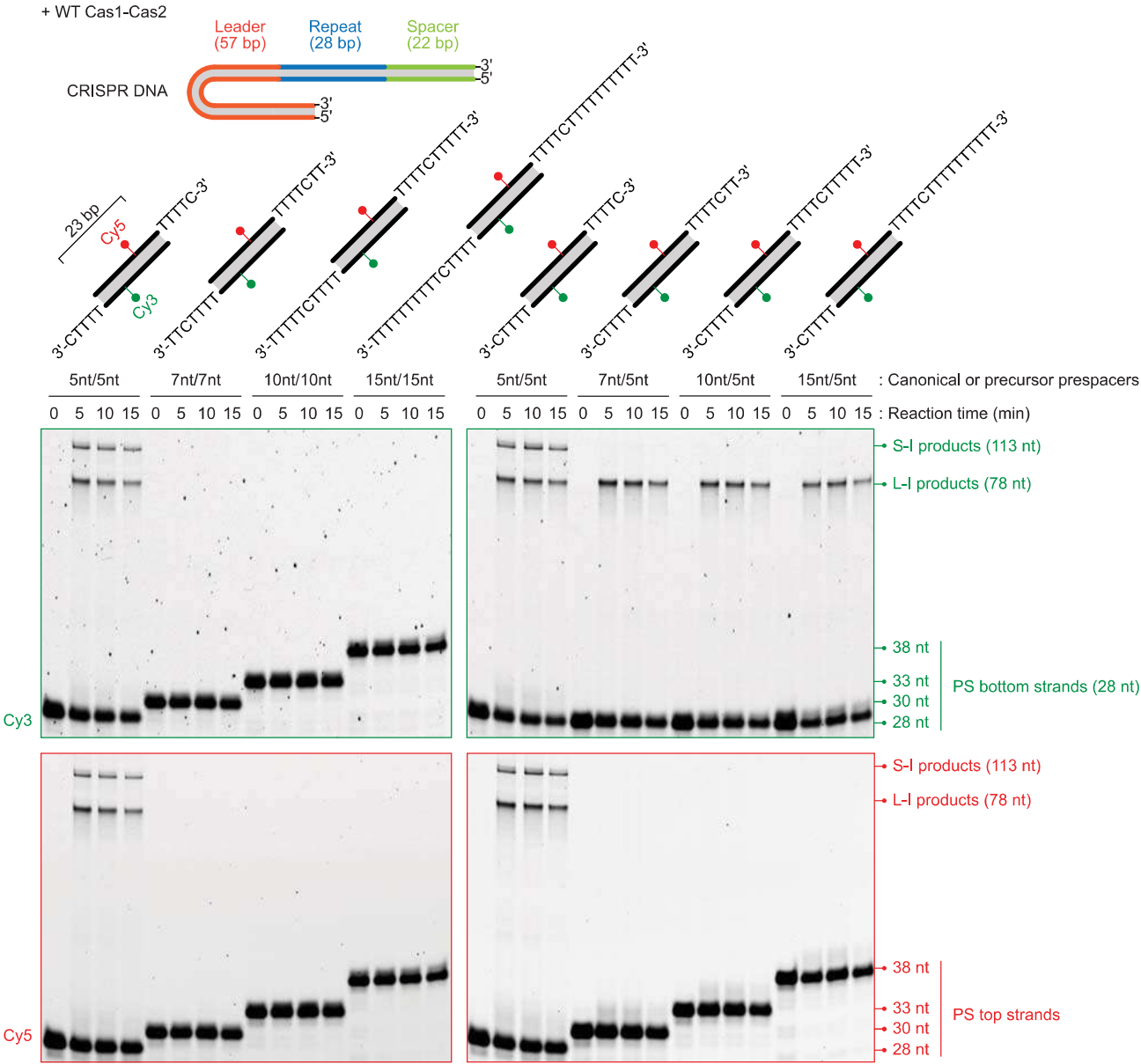
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | PAM-dependent ssDNA capture by Cas1–Cas2.

a, Representative time traces of a single Cas1–Cas2 complex binding to ssDNAs that contain different numbers of PAM sites. DNA was added at $t = 5$ s. The insets show a snapshot of the field of view taken after a 10-min incubation.

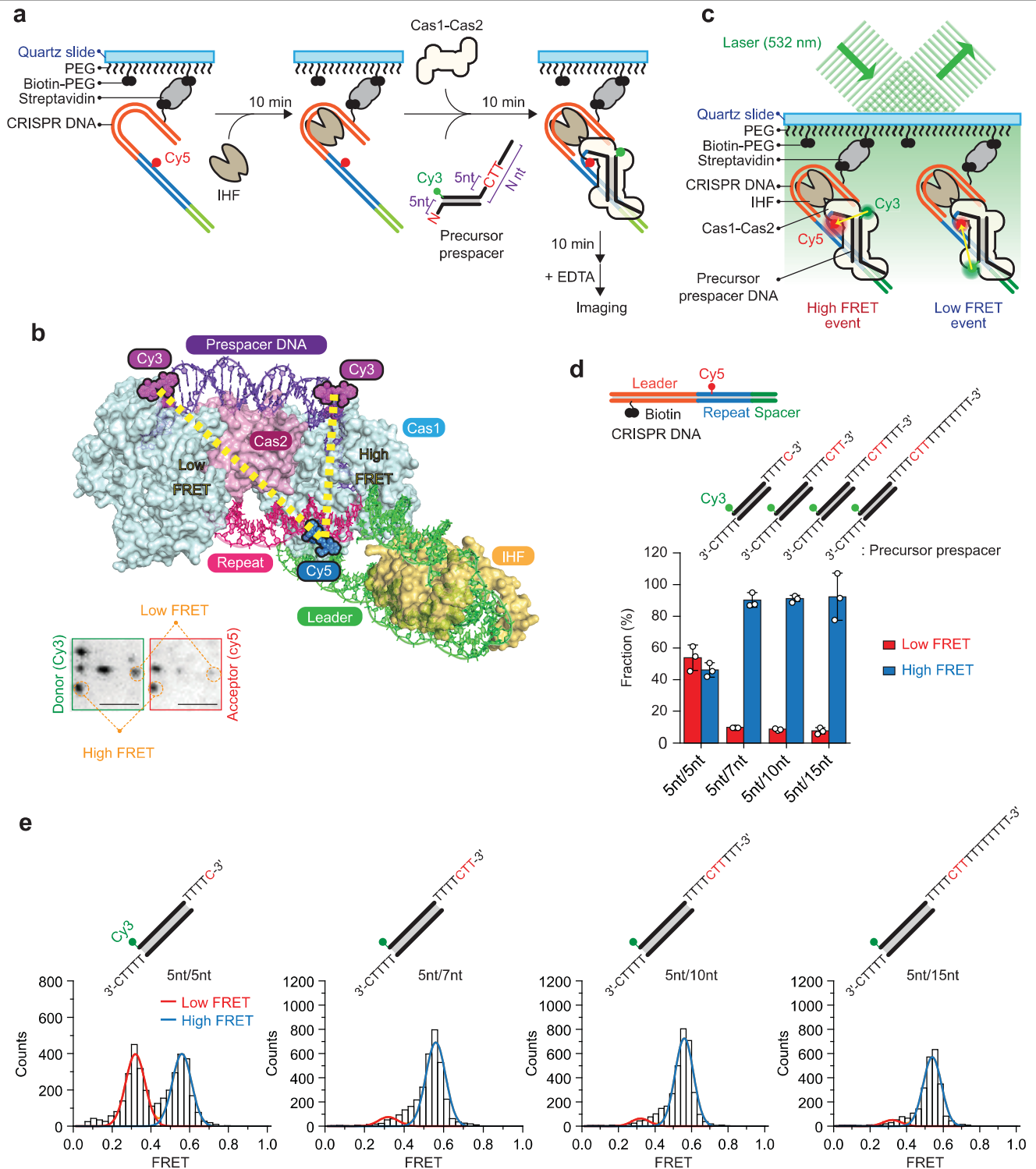
b, Cumulative probability of the arrival times with a single-exponential fit (solid line) that was used to determine the binding frequency (k_{on}). Data are mean \pm 95% CI, obtained by bootstrap analysis from a single replicate with $n \geq 300$ individual molecules. **c**, Dwell-time distributions of binding events for ssDNAs that contain different numbers of PAM sites. Average dwell times (τ_{off}) are mean \pm 95% CI, obtained by bootstrap analysis of a single replicate with $n \geq 500$ individual molecules. **d**, k_{on} of ssDNA substrates of various lengths, containing one PAM site. Data are mean \pm 95% CI, obtained by bootstrap analysis from a single replicate with $n \geq 300$ individual molecules. **e**, k_{off} of ssDNA substrates of various lengths, containing one PAM site. Data are mean \pm 95% CI, obtained by bootstrap analysis of a single replicate with $n \geq 500$

individual molecules. **f**, Model of a facilitated diffusion mechanism for PAM-dependent ssDNA binding by Cas1–Cas2. Cas1–Cas2 binds a non-specific (non-PAM) region on ssDNA, which is followed by rapid facilitated diffusion and PAM recognition. Although the diffusive movement cannot be directly observed with the time resolution of our single-molecule assay (0.1 s), the effects can be seen when the ssDNA substrate is extended (see **d**). When the length is increased, the measured binding frequency increases (commonly referred to as the antenna effect), which suggests that Cas1–Cas2 uses facilitated diffusion to locate PAM sequences. **g**, EMSAs on various ssDNA and dsDNA substrates with increasing amounts of wild type Cas1–Cas2. Top, EMSAs with ssDNAs without (Cy3) or with (Cy5) a PAM sequence. Bottom, EMSAs with a precursor prespacer DNA substrate that is annealed, or with two ssDNAs added simultaneously. The bands that correspond to the bound and unbound fractions are indicated on the right. For gel source data, see Supplementary Fig. 1. Data are representative of three replicates with similar results (**a–e**, **g**).



Extended Data Fig. 4 | In vitro integration assay with precursor prespacer DNAs with 3'-overhangs of various lengths. In vitro integration assay using a linear CRISPR DNA and canonical or precursor prespacer DNAs with 3'-overhangs of different lengths. Full-site integration of a mature prespacer DNA (28 nt) results in a 78-nt leader-side integration (L-I) product and a 113-nt spacer-side integration (S-I) product. The top and bottom strands of the canonical or precursor prespacer DNA substrates were labelled with Cy5 and

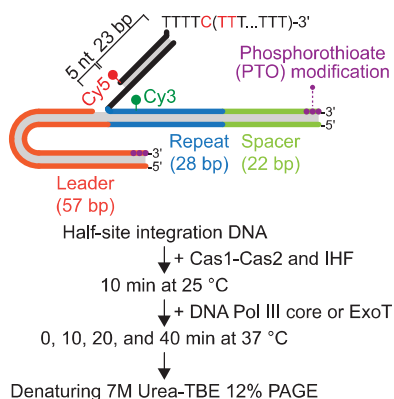
Cy3, respectively. Samples were run on a 7 M urea denaturing 20% TBE-PAGE, after which images were collected with a Typhoon scanner. Only those precursor prespacers with the canonical size of 5 nt (5'-TTTTC-3') in the 3'-overhang(s) were efficiently incorporated into the CRISPR locus to yield leader-side (spacer-side) integration products. For gel source data, see Supplementary Fig. 1. Data are representative of three replicates with similar results.



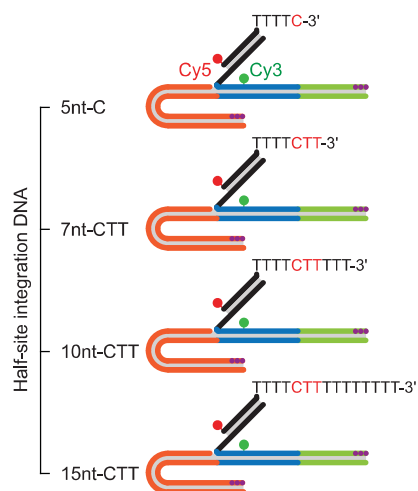
Extended Data Fig. 6 | Single-molecule leader-side integration assay for the PAM-deficient end of asymmetrically trimmed precursor prespacer DNAs.

a, Schematic of the single-molecule FRET assay that was used to observe the orientation of integrated spacers. Biotinylated CRISPR DNA was labelled with Cy5 in the repeat region (5 nt away from the leader-repeat junction). Precursor prespacer DNA was labelled with Cy3 at the 5'-end of the top strand. **b**, Expected FRET from the single-molecule assay based on structural modelling (PDB: 5WFE). Representative CCD images in donor (green box) and acceptor (red box) channels are included as insets and indicated with representative high and low FRET states. **c**, smFRET design for assessing the orientation of integrated

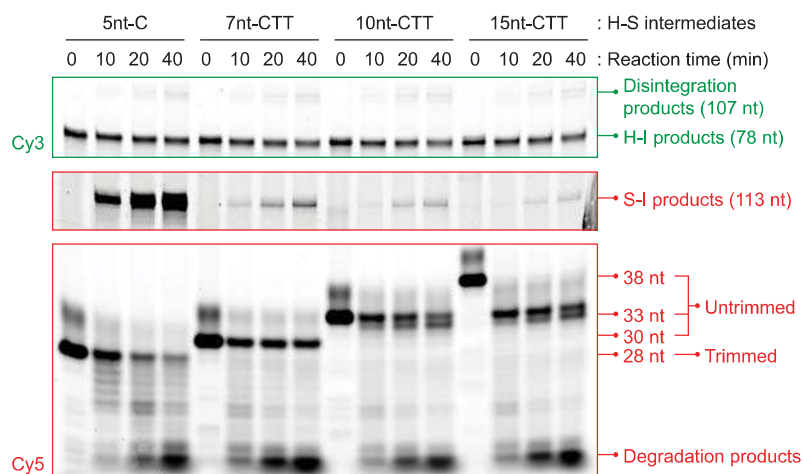
products. The 5'-end of the top strand was labelled with Cy3. Integration of the 3'-end of the bottom strand at the leader side exhibits high FRET, and integration of the 3'-end of the top strand shows low FRET. **d**, Fractions of high- and low-FRET events after the integration reaction. Data are mean \pm s.e.m. from three independent measurements ($n=3$) with $n \geq 3,000$ molecules for each measurement. **e**, FRET efficiency histograms of precursor prespacer DNAs with a 3'-overhang length that is optimal (28-nt) for the PAM-deficient strand and non-optimal for the PAM-containing strand. Solid lines represent Gaussian fits to obtain the high and low FRET populations. Data are representative of three replicates with similar results.

a**b**

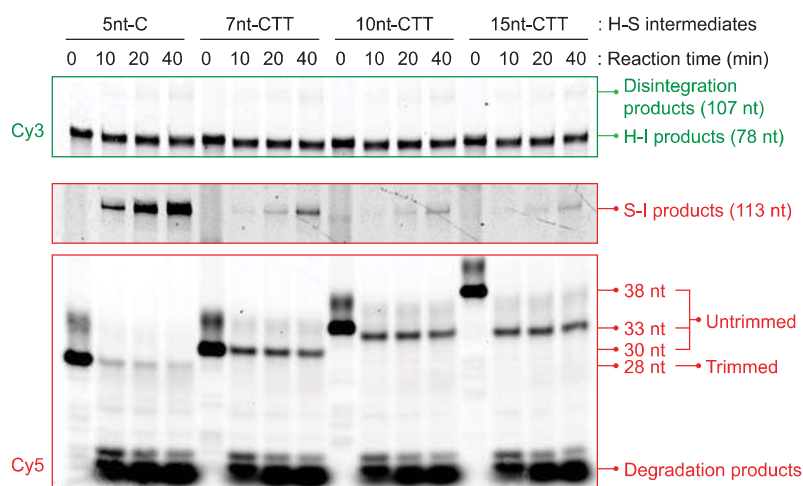
+ WT Cas1-Cas2
+ DNA Pol III core or ExoT

**c**

+ DNA Pol III core

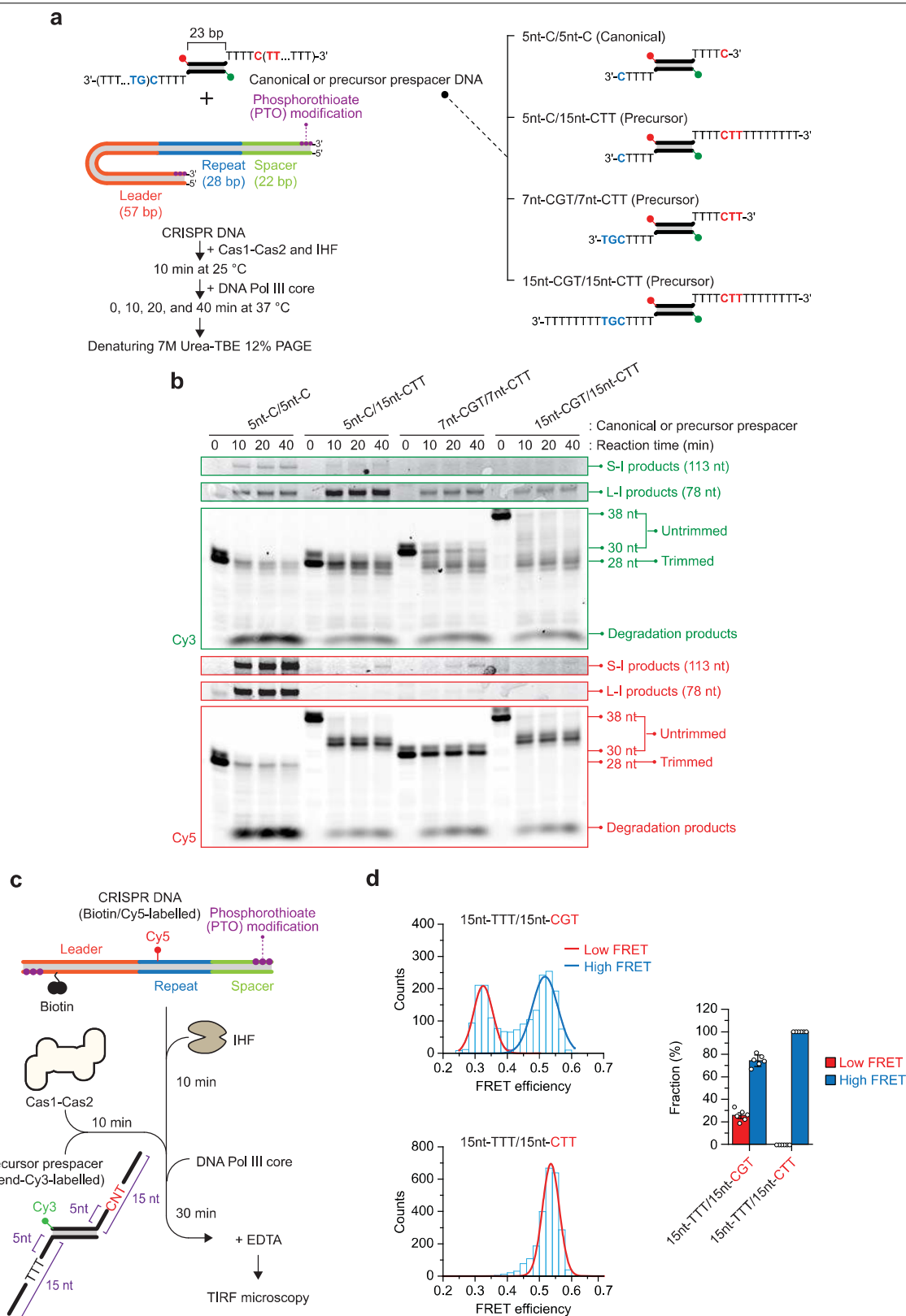
**d**

+ ExoT



Extended Data Fig. 7 | Maturation and integration of the PAM-containing end in half-site intermediates. **a**, Design for in vitro and single-molecule trimming-driven integration assays. The last three backbone phosphodiester bonds from the 3'-end of CRISPR DNA were modified with PTO (purple) to prevent degradation by 3'-5' exonucleases. **b**, Schematic of the substrates used for the in vitro trimming-driven full-site integration assay. **c**, **d**, Gel images from the in vitro trimming-driven full-site integration assay with DNA Pol III core (**c**)

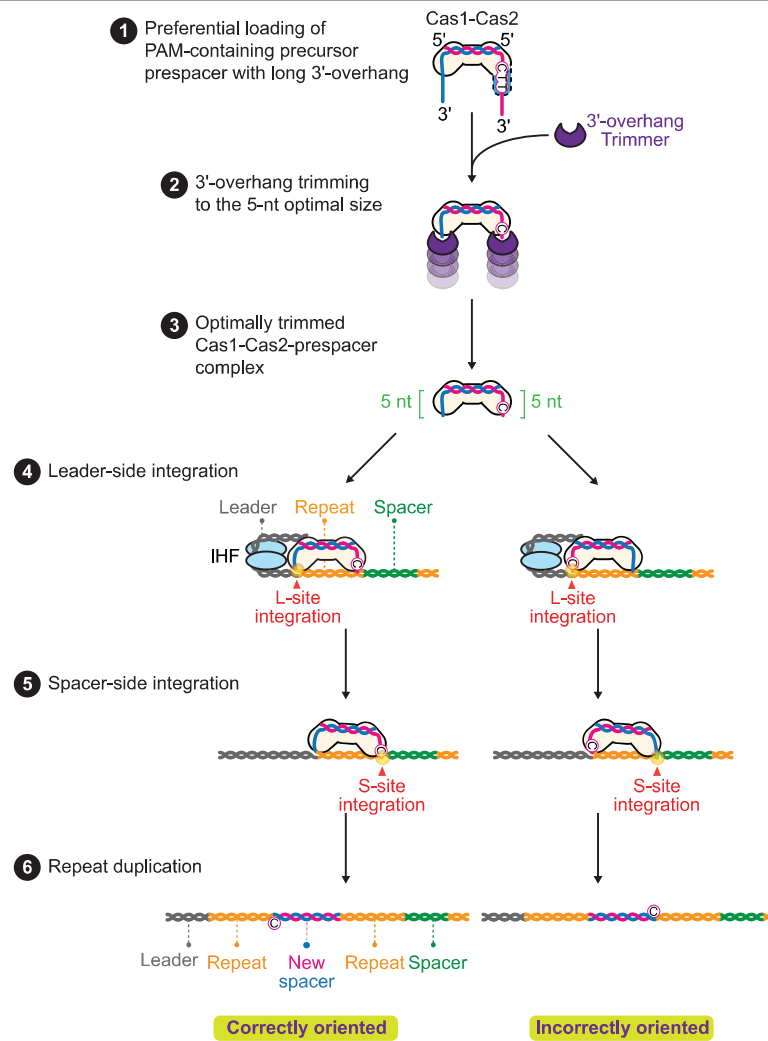
or ExoT (**d**). Unreacted half-site (H-S) intermediates and disintegrated products are shown in the Cy3 image. Spacer-side integration products and processed top prespacer strands are shown in the Cy5 images. For clarity, the bottom part (below 50 nt) of the Cy5 image was separated from the top part (around 70–130 nt) and adjusted with a different contrast. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 8 | Reconstitution of trimming-driven integration.

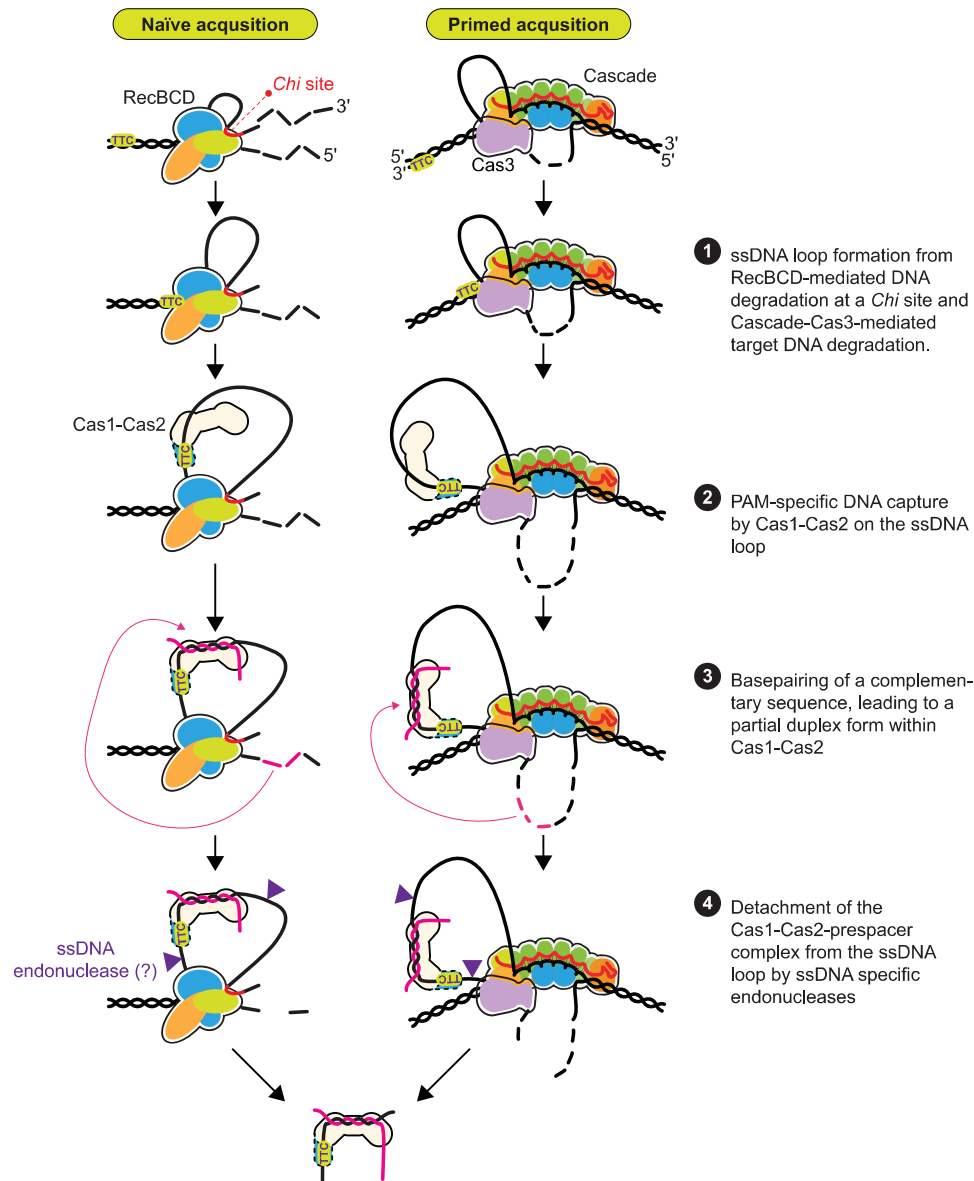
a, Design of in vitro and single-molecule trimming-driven integration assays. The last three backbone phosphodiester bonds from the 3'-end of CRISPR DNA were modified with PTO (purple) to prevent degradation by 3'-5' exonucleases. **b**, Representative gel images from the in vitro trimming-driven integration assay with DNA Pol III core. The contrast of areas of spacer-side and leader-side integration products was adjusted for optimal visibility. For gel source data, see Supplementary Fig. 1. **c**, **d**, Single-molecule assay for biased integration of

PAM-containing precursor prespacer DNA. **c**, Schematic of the experimental procedure used for trimming-driven integration assays at the single-molecule level. **d**, FRET efficiency histograms from the trimming-driven integration assay. The Cy3-labelled top strand of precursor prespacer DNAs had either CGT (non-PAM) or CTT (PAM). Solid lines represent Gaussian fits to obtain the high and low FRET populations. The bar plot displays fractions of high and low FRET populations after the integration reaction. Data are presented as mean \pm s.e.m. ($n = 6$). Data are representative of three replicates with similar results.



Extended Data Fig. 9 | Previous unbiased trimming model of spacer acquisition. In this model, the 3'-overhangs of PAM-containing Cas1-Cas2-selected prespacer precursors are trimmed symmetrically to the canonical length of 5 nt. The PAM-derived 3'-end of trimmed prespacer can be integrated

either in the leader side or the spacer side of the CRISPR DNA. Therefore, the probability of a correctly oriented spacer in the CRISPR DNA is 50% (which does not agree with spacer acquisition in vivo).



Extended Data Fig. 10 | Biased trimming model of spacer acquisition. Our model suggests that ssDNA loops generated from RecBCD- and Cascade-Cas3-mediated DNA degradation provide the initiation sites for Cas1-Cas2 docking and prespacer selection.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|--|
| Data collection | Single-molecule data was collected using a modified version of "SINGLE" software that was developed by the lab of Prof. Dr. Ha. The original version is available at: https://cplc.illinois.edu/software/ . Gel images were collected using the Typhoon imager (GE Healthcare) and its corresponding softwares (Amersham Typhoon version 1.1.0.7 and ImageQuant TL Version 8.1). |
| Data analysis | Single-molecule data were first analyzed using custom IDL scripts that allowed automatic extraction of single-molecule time trajectories. Single-molecule trajectories were analyzed using custom MATLAB scripts that allowed manual selection and processing of the trajectories. The scripts are available upon request. Statistical analysis was performed using both standard functions in IDL (version 8.2), MATLAB R2017b and OriginPro 2015 (version Sr1 b9.2.257). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

We have provided a full data availability statement in the manuscript. No data with mandated deposition.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Experiments that yielded a single value per data point were performed as three individual replicates. In contrast, experiments that yielded large numbers of data points (>500) were repeated once. The error of these large datasets was obtained through bootstrap analysis which allows assigning measures of accuracy (defined in terms of bias, variance, confidence intervals, prediction error or some other such measure) to sample estimates. Bootstrapped data represents mean \pm 95% confidence interval (CI) were obtained from a single replicate with $n \geq 500$ individual molecules. The minimal number of data points for bootstrap analysis is ~ 50 . We doubled the minimum number of data points to ensure accuracy.
Data exclusions	For the single-molecule analysis of DNA binding, all traces were selected and analyzed that exhibited FRET. In contrast, traces that displayed donor-only sign were excluded from the analysis as these molecules may correspond to single-stranded DNA molecules in solution. In some cases attempts to reproduce data on microscope slides of low quality (poor surface-passivation) failed. These datasets were excluded and repeated on high-quality microscope slides.
Replication	To verify the reproducibility, three individual experiments were performed and analyzed. We hereby confirm we can reproduce the data presented in this manuscript. All presented data is representative of three replicates that yielded similar results.
Randomization	To allow accurate comparison between the data presented in the figures, the substrates that were used for each figure were grouped and the data was obtained at the same day. Randomization is not applicable in this case.
Blinding	Analysis performed in this manuscript were not blinded. However, strict selection criteria were used to ensure reproducibility of the analysis. Moreover, the analysis of the data was consistent between two individuals confirming the consistency and reproducibility of the analysis method.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging


Structural basis of energy transfer in *Porphyridium purpureum* phycobilisome

<https://doi.org/10.1038/s41586-020-2020-7>

Received: 15 July 2019

Accepted: 19 December 2019

Published online: 19 February 2020

 Check for updates

Jianfei Ma^{1,3}, Xin You^{1,3}, Shan Sun^{1,3}, Xiaoxiao Wang², Song Qin² & Sen-Fang Sui^{1✉}

Photosynthetic organisms have developed various light-harvesting systems to adapt to their environments¹. Phycobilisomes are large light-harvesting protein complexes found in cyanobacteria and red algae^{2–4}, although how the energies of the chromophores within these complexes are modulated by their environment is unclear. Here we report the cryo-electron microscopy structure of a 14.7-megadalton phycobilisome with a hemiellipsoidal shape from the red alga *Porphyridium purpureum*. Within this complex we determine the structures of 706 protein subunits, including 528 phycoerythrin, 72 phycocyanin, 46 allophycocyanin and 60 linker proteins. In addition, 1,598 chromophores are resolved comprising 1,430 phycoerythrobilin, 48 phycourobilin and 120 phycocyanobilin molecules. The markedly improved resolution of our structure compared with that of the phycobilisome of *Griffithsia pacifica*⁵ enabled us to build an accurate atomic model of the *P. purpureum* phycobilisome system. The model reveals how the linker proteins affect the microenvironment of the chromophores, and suggests that interactions of the aromatic amino acids of the linker proteins with the chromophores may be a key factor in fine-tuning the energy states of the chromophores to ensure the efficient unidirectional transfer of energy.

Light absorption is the first step of photosynthesis. The membrane-extrinsic soluble phycobilisomes (PBSs) are responsible for the majority of light capture in cyanobacteria and red algae^{2–4}. PBSs are composed of phycobiliproteins (PBPs) and linker proteins², and sunlight is absorbed and the energy transferred by open-chain tetrapyrrole chromophores that covalently bind to PBPs and some linker proteins^{2,6}. A heterodimer of two different PBP subunits (α - and β subunits) assembles into a ring-shaped ($\alpha\beta$)₃ trimer, which serves as the basic unit for the PBS assembly². The typical PBS consists of several peripheral rods surrounding the central core^{1,7}. Solar photonic energy absorbed by the peripheral rods can be rapidly funnelled to the core¹ and eventually to the terminal emitters—chromophores of the core—membrane linker protein (L_{CM})^{8–10} or allophycocyanin D (ApcD)^{11,12}—then transferred to photosystems I and II^{2,10,13,14}.

Four morphological types of PBS have been observed: hemidisoidal^{15–20}, hemiellipsoidal^{21,22}, block-type²³ and bundle-type²⁴. We recently solved the structure of the block-shaped PBS from the red alga *G. pacifica* at 3.5 Å resolution, which provided the detailed architecture of the intact PBS⁵. However, to our knowledge there have been no high-resolution structures reported for other morphological types of PBS. Moreover, although we determined the locations of all chromophores of the *G. pacifica* PBS, owing to resolution limitations we could not reveal how the energies of the bilins are modulated by their surroundings. To address these questions, we resolved the structure of a typical hemiellipsoidal-shaped PBS^{21,22,25} from *Porphyridium purpureum*, one of the few unicellular red algae and a widely used model alga^{26–28}. The resulting structure, determined by cryo-electron microscopy

(cryo-EM) at 2.82 Å resolution, reveals how the linker proteins affect the microenvironments of chromophores.

Overall structure

Intact PBSs were purified from *P. purpureum* and their subunit composition and spectroscopic properties are shown in Extended Data Fig. 1. The overall resolution of the resulting structure is 2.82 Å, with a higher resolution of 2.68 Å for the core region (Fig. 1a, Extended Data Fig. 2, Extended Data Table 1). Some long loops in L_{CM} , which are absent in the electron microscopy map of the *G. pacifica* PBS, are clearly resolved in this reconstruction (Fig. 1b). The different types of bilins can be unambiguously assigned on the basis of the densities and the dihedral angles in combination with the results of published biochemical analysis^{29–35} that classified the phycocyanin in *P. purpureum* as R-phycocyanin, containing one phycoerythrobilin (PEB) and one phycocyanobilin (PCB) on the β subunit³⁴ (Fig. 1c, Extended Data Fig. 3). In total, we built 706 protein subunits comprising 528 phycoerythrin subunits, 72 phycocyanin subunits, 46 allophycocyanin subunits and 60 linker proteins, and we assigned 1,598 chromophores (Extended Data Table 2a).

The two-fold symmetric PBS resembles an opened fan from the face view and has an oval outline from the top view, with approximate dimensions of 610 Å (length), 390 Å (height) and 380 Å (thickness) (Fig. 1d). It contains 14 peripheral rods surrounding a pyramidal-shaped core (Extended Data Fig. 4a). The core contains one top cylinder (B) formed by two allophycocyanin trimers stacked back to back, and two bottom

¹State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing, China. ²Yantai Institute of Coast Zone Research, Chinese Academy of Sciences, Yantai, China. ³These authors contributed equally: Jianfei Ma, Xin You, Shan Sun. ✉e-mail: suisf@mail.tsinghua.edu.cn

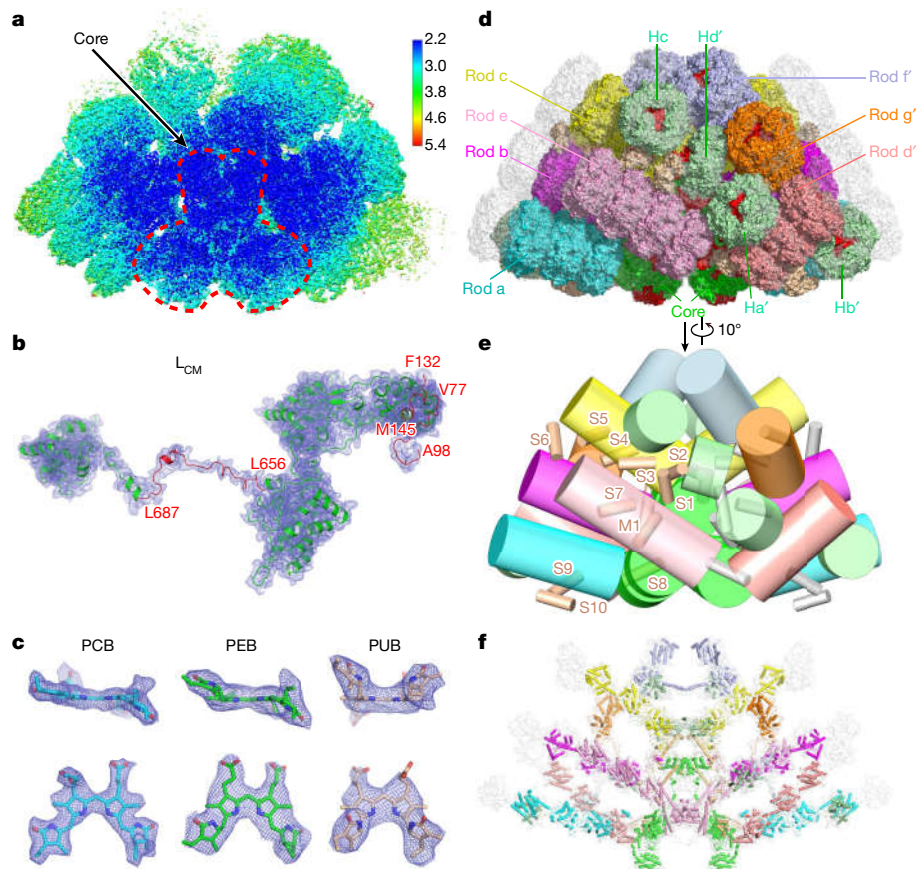


Fig. 1 | Overall architecture of the PBS from *P. purpureum*. **a**, Local resolution map of the PBS from *P. purpureum*. The map was estimated with ResMap and generated in Chimera. **b**, The density (mesh) for the linker protein L_{CM} superimposed with its atomic model (cartoon). Three loops (red)—V77 to A98, F132 to M145 and L656 to L687—are clearly resolved. **c**, The densities (mesh) of the representative PCB, PEB and phycourobilins (PUB) bilins (stick representation) show their different coplanarities. **d**, Overall structure of the PBS shown in surface representation. The rods are shown in different colours, and the core is coloured green. All extra hexamers, individual monomers and β

subunits, and linker proteins are coloured light green, wheat and red, respectively. The superimposed structure of the *G. pacifica* PBS is coloured grey. **e**, Schematic showing the distribution of the individual phycoerythrin ($\alpha\beta$) monomers and phycoerythrin β subunits. One half (M1 and S1–S10) are coloured wheat and the other half are grey. **f**, Structures of all linker proteins shown in cartoon representation, from the same view as in **d**. They are the same colours as the hexamers in which they are located. The superimposed structures of linker proteins of the *G. pacifica* PBS are coloured grey.

cylinders (A and A'), each of which is assembled by one ($\alpha\beta$)₆ allophycocyanin hexamer and one allophycocyanin trimer (Extended Data Fig. 4b). The rods are categorized into two types according to their PBP composition: type I rods (Ra/Ra', Rb/Rb' and Rc/Rc') are composed of both phycoerythrin and phycocyanin, whereas type II rods (Rd/Rd', Re/Re', Rf/Rf' and Rg/Rg') are composed entirely of phycoerythrin (Extended Data Fig. 4d). Except for rods Rf/Rf' and Rg/Rg', each of which contains two phycoerythrin hexamers in both PBs, the number of phycoerythrin hexamers in each of the remaining rods is one fewer in the *P. purpureum* PBS than in the *G. pacifica* PBS (Extended Data Fig. 4c, d). The *P. purpureum* PBS also contains extra phycoerythrin hexamers; however, the hexamers He/He', which are located near the surface of the outermost hexamer of rods Rb/Rb' and Rc/Rc' in the *G. pacifica* PBS, are absent owing to the short lengths of these rods in the *P. purpureum* PBS (Fig. 1d, Extended Data Fig. 4c). There are 2 individual phycoerythrin ($\alpha\beta$) monomers M1 (M1') and 20 individual phycoerythrin β subunits S1–S10 (S1'–S10') interspersed throughout the whole PBS (Fig. 1e). These components fill the empty spaces outside the rods, core and extra hexamers, and may stabilize the PBS. The *P. purpureum* PBS is aligned well with the *G. pacifica* PBS, but has a smaller size owing to the reduced number of phycoerythrin hexamers, indicating a similar overall organization of the rods and core (Fig. 1d, Extended Data Fig. 4c). The molecular mass of the *P. purpureum* PBS is

14.7 MDa, which is less than that of the *G. pacifica* PBS (18.0 MDa) after considering the molecular masses of the chromophores.

The skeleton formed by the linker proteins is very similar in the PBs from both *P. purpureum* and *G. pacifica* (Fig. 1f), and both contain 17 types of linker protein. Superimposing the two PBs indicate that they share very high structural conservation, except for the rod linker protein L_{R6} (Extended Data Fig. 4e). L_{R6} of the *P. purpureum* PBS contains the Pfam00427 domain, instead of the CBDy domain that is present in L_{R6} of the *G. pacifica* PBS; this is in accordance with the overall tendency for the *P. purpureum* PBS to contain fewer bilins than the *G. pacifica* PBS. The roles of linker proteins in the assembly of the PBS—such as the sequential interactions between them to form the skeleton, the extensive contacts between them and the hexamers, and the α -helix-mediated interactions between L_{RC} proteins and the core—are common between these two PBs, highlighting their evolutionary conservation (Extended Data Fig. 5).

There are 120 PCBs, 1,430 PEBs and 48 phycourobilins in *P. purpureum* PBS (Extended Data Table 2a). The phycourobilin content in the *P. purpureum* PBS is considerably lower than that in the *G. pacifica* PBS; this is because all phycourobilins in the *P. purpureum* PBS originate solely from the $L_{R\gamma}$ proteins, whereas in the *G. pacifica* PBS—besides the $L_{R\gamma}$ proteins—all phycoerythrin β subunits also contain phycourobilin³⁶. The lower phycourobilin content and the reduced number of total bilins

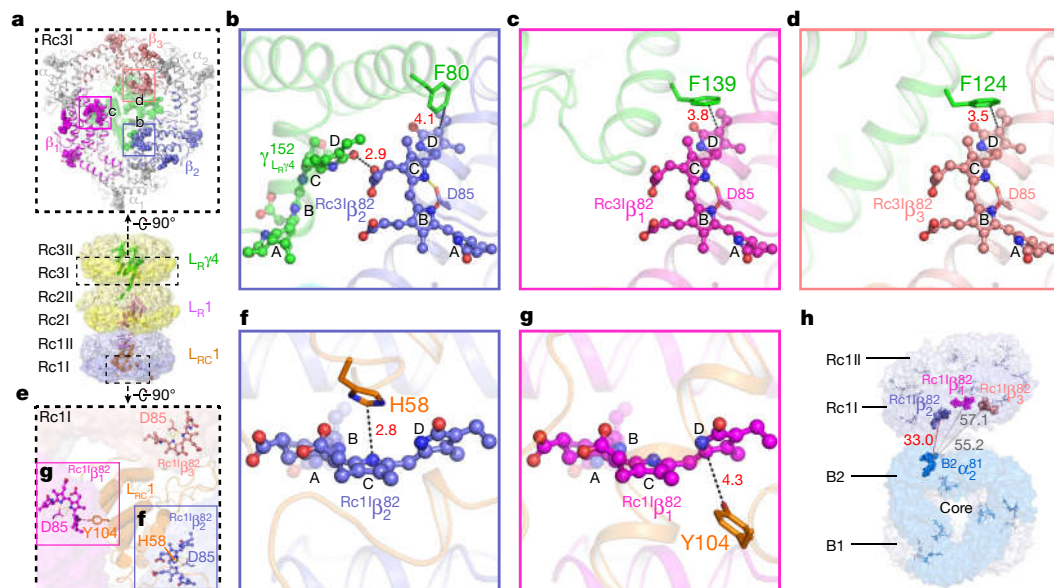


Fig. 2 | Interactions of the linker proteins L_{Ry} and L_{Rc} with chromophores in the rod Rc. **a**, Bottom, overall structure of the rod Rc with the hexamers shown in surface representation and the linker proteins shown in cartoon representation. Top, structure of the layer Rc3I. Proteins and bilins are shown in cartoon and sphere representations, respectively. Three β subunits are coloured differently and the β 82 PEBs are boxed and analysed in detail in **b–d**. **b**, The interactions between the residue F80 and the bilin $Y_{L_{Ry}4}^{152}$ from $L_{Ry}4$ with the bilin $Rc3I\beta_2^{82}$. **c**, The interaction between F139 from $L_{Ry}4$ and the bilin $Rc3I\beta_1^{82}$. **d**, The interaction between F124 from $L_{Ry}4$ and the bilin $Rc3I\beta_3^{82}$. **e**, A focused view of the structure of the layer Rc1I showing the central triangle area. PBPs, the linker protein, bilins and residues are shown in surface, cartoon, ball-and-stick and stick representations, respectively. Two β 82 PCBs are boxed and analysed in detail in **f** and **g**. **f**, The strong interaction between H58 from $L_{Rc}1$ and the bilin $Rc1I\beta_2^{82}$. **g**, The interaction between Y104 from $L_{Rc}1$ and the bilin $Rc1I\beta_1^{82}$. **h**, $Rc1I\beta_2^{82}$ has the shortest distance (33.0 Å, red line) between the rod and the core compared to $Rc1I\beta_1^{82}$ (57.1 Å, grey line) and $Rc1I\beta_3^{82}$ (55.2 Å, grey line).

d, The interaction between F124 from $L_{Ry}4$ and the bilin $Rc3I\beta_3^{82}$. **e**, A focused view of the structure of the layer Rc1I showing the central triangle area. PBPs, the linker protein, bilins and residues are shown in surface, cartoon, ball-and-stick and stick representations, respectively. Two β 82 PCBs are boxed and analysed in detail in **f** and **g**. **f**, The strong interaction between H58 from $L_{Rc}1$ and the bilin $Rc1I\beta_2^{82}$. **g**, The interaction between Y104 from $L_{Rc}1$ and the bilin $Rc1I\beta_1^{82}$. **h**, $Rc1I\beta_2^{82}$ has the shortest distance (33.0 Å, red line) between the rod and the core compared to $Rc1I\beta_1^{82}$ (57.1 Å, grey line) and $Rc1I\beta_3^{82}$ (55.2 Å, grey line).

in *P. purpureum* PBS are consistent with the fact that *P. purpureum* live at the sea surface³⁷, where the light intensity is higher compared with that beneath the sea surface where *G. pacifica* are found³⁷.

Interactions of L_{Ry} s with chromophores

The $(\alpha\beta)_3$ trimers of the phycocyanins, phycoerythrins and allophycocyanins have very similar ring-like structures, in which the central cavity is a common feature. Three β 82 chromophores are located near to the inner cavity and are subjected to interactions with linker proteins³⁸ (Fig. 2a). The trimers Rc3I and Rd3I—from the type I rod Rc and the type II rod Rd, respectively—are used here to illustrate how the rod linkers $L_{Ry}4$ and $L_{Ry}5$ interact with the β 82 chromophores (Fig. 2). Each of the β 82 PEBs of Rc3I (denoted $Rc3I\beta_1^{82}$, $Rc3I\beta_2^{82}$ and $Rc3I\beta_3^{82}$) is bound by two hydrogen bonds formed between the nitrogen atoms of the pyrrole rings B and C and the carboxyl group of the D85 residue of the β subunit (Fig. 2b–d), in agreement with the crystal structure of R-phycoerythrin³⁹. In particular, three aromatic residues of $L_{Ry}4$ (F80, F139 and F124) are located close to rings D of $Rc3I\beta_1^{82}$, $Rc3I\beta_2^{82}$ and $Rc3I\beta_3^{82}$, respectively, which will stabilize ring D and may expand the conjugation of the system owing to π – π interactions (Fig. 2b–d). Moreover, an extra PEB from $L_{Ry}4$ ($Y_{L_{Ry}4}^{152}$) is adjacent—and very close—to the chromophore $Rc3I\beta_2^{82}$: the distance between their nearest two atoms is only 2.9 Å (Fig. 2b). Therefore, the chromophore pair may further downgrade the energy level of $Rc3I\beta_2^{82}$ owing to excited-state coupling^{39,40}, with the result that $Rc3I\beta_2^{82}$ is probably at a lower energy level compared with that of $Rc3I\beta_1^{82}$ and $Rc3I\beta_3^{82}$. Notably, the trimer Rd3I in the type II rod Rd displays similar structural features: one aromatic residue is close to each of the β 82 PEBs, and an additional bilin from the linker $L_{Ry}5$ ($Y_{L_{Ry}5}^{135}$) also resides close to the $Rd3I\beta_2^{82}$ (Extended Data Fig. 6a–d). Moreover, structural superimposition reveals that the L_{Ry} linker proteins in the outmost hexamers of various rods of PBSs from both *P. purpureum* and *G. pacifica* also have similar structures (Extended Data Fig. 7a, b). These key aromatic residues, and the cysteine residues that are used to link

the bilins, are well conserved in red algae (Extended Data Fig. 7c). The structural features of the interaction between L_{Ry} and β 82 PEBs therefore suggest that β_2^{82} PEB is in the lowest energy state among the three β 82 PEBs, and that energy migration through β_2^{82} PEB could be the major route by which to pass energy through the rod.

Interactions of L_{Rc} s with chromophores

Energy is then transferred along the rods to the triangular area of the core-proximal hexamer (Fig. 2e). Two types of rod use different rod–core linkers to associate with the core. For Rc—which uses the linker $L_{Rc}1$ —a heterocyclic residue (H58) from $L_{Rc}1$ is located close to $Rc1I\beta_2^{82}$, with a minimum distance of 2.8 Å (Fig. 2f). The pyrrole group of H58 can form a strong π – π interaction with rings B and C of $Rc1I\beta_2^{82}$ (ref. 41). However, just one aromatic residue (Y104) from $L_{Rc}1$ forms a relatively weak π – π interaction with ring D of $Rc1I\beta_1^{82}$ —because of the longer distance (4.3 Å) compared with that between H58 and $Rc1I\beta_2^{82}$ —and no aromatic residues interact with $Rc1I\beta_3^{82}$ (Fig. 2e, g). Therefore, modified by the specific surroundings, $Rc1I\beta_2^{82}$ may be in the lowest-energy state among the three β 82 chromophores. Moreover, $Rc1I\beta_2^{82}$ has the shortest distance to the core compared with $Rc1I\beta_1^{82}$ and $Rc1I\beta_3^{82}$ (Fig. 2h), which further suggests that it may act as an energy-transit station, converging the energy absorbed by the rod and transferring it to the core. Similar situations are found for another two rods of type I and the three type I rods in *G. pacifica* (Extended Data Fig. 7d). This histidine residue is conserved completely across different red algal and cyanobacterial species, which is indicative of its functional importance (Extended Data Fig. 7e).

For Rd—which uses $L_{Rc}2$ —two aromatic residues from $L_{Rc}2$ form parallel-displaced and T-shaped π – π interactions with $Rd1I\beta_2^{82}$, respectively (Extended Data Fig. 6f). By contrast, each of another two β 82 PEBs interacts with only one aromatic residue (Extended Data Fig. 6g, h). Superimposition of the L_{Rc} proteins ($L_{Rc}2$ and $L_{Rc}3$) of type II rods from both *P. purpureum* and *G. pacifica* shows that two such aromatic

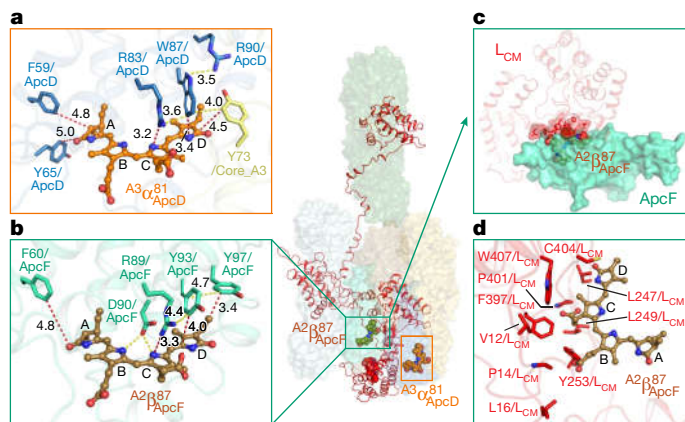


Fig. 3 | The bilins of ApcD and ApcF and their surrounding residues. **a**, The surrounding residues of the bilin $A^3\alpha_{ApcD}^{81}$. The residues and bilin are shown in stick and ball-and-stick representations, respectively. **b**, The interactions between $A^2\beta_{ApcF}^{87}$ and ApcF. The residues and bilin are shown in stick and ball-and-stick representations, respectively. **c**, $A^2\beta_{ApcF}^{87}$ is buried by a contiguous hydrophobic cap formed by the linker protein L_{CM} . L_{CM} is shown in cartoon representation in red and the cap is displayed in surface representation. $A^2\beta_{ApcF}^{87}$ is shown in sphere representation. **d**, The interactions between the cap and $A^2\beta_{ApcF}^{87}$. The hydrophobic residues in the cap are shown in stick representation in red.

residues exist in all cases (Extended Data Fig. 7f). Sequence alignment also shows that these aromatic residues are conserved, which suggests their importance in fine-tuning the energies of the bilins (Extended Data Fig. 7g).

Key chromophores in the core

Previous studies have shown that three PCB chromophores in ApcD, ApcF and the α subunit of L_{CM} ($A^3\alpha_{ApcD}^{81}$, $A^2\beta_{ApcF}^{87}$ and $A^2\alpha_{L_{CM}}^{186}$) perform critical functions in energy transfer in the core. However, why each of these chromophores have unique functions remains to be clarified. Here we use our high-resolution structure to analyse the immediate surroundings of these key core chromophores in their native states.

Functionally, ApcD is the main protein responsible for energy transfer to photosystem I¹³. In our structure, two aromatic residues—W87 from ApcD and Y73 from the β subunit of the core trimer A3—form T-shaped and parallel-displaced π – π interactions with $A^3\alpha_{ApcD}^{81}$, respectively; this enhances the tight fitting of ring D (Fig. 3a), which is consistent with the crystal structure of ApcD from *Synechocystis* PCC 6803 (PDB: 4PO5)¹². Notably, we observed that W87 was surrounded by R83 and R90 from ApcD and Y73 from the core A3, which provided two cation– π interactions and one T-shaped π – π interaction to W87, respectively (Fig. 3a). It can therefore be inferred that the presence of these three residues is necessary to stabilize the orientation of W87, which is critical for the conformation of $A^3\alpha_{ApcD}^{81}$. In addition, more π -related interactions between residues and $A^3\alpha_{ApcD}^{81}$ are extracted from the high-resolution structure of the entire PBS. The cationic side chain of R83 extends to the top of ring C of $A^3\alpha_{ApcD}^{81}$, forming the cation– π interaction⁴². F59 and Y65 may contribute two additional π – π interactions to ring A of $A^3\alpha_{ApcD}^{81}$ (Fig. 3a). We then superimposed ApcDs from *P. purpureum*, *G. pacifica* and *Synechocystis* PCC 6803 with the α subunit of the core A3, and found that W87, F59 and Y65 are common to all ApcD molecules, although Y65 was replaced by Y66 in the A3 α subunit (Extended Data Fig. 8a).

ApcF plays a crucial role in energy migration to the terminal chromophore of L_{CM} ^{14,43}. Analysis of the PCB pocket of ApcF showed that the positively charged R89 formed one cation– π interaction with ring C of $A^2\beta_{ApcF}^{87}$ and Y93 and Y97 formed one T-shaped π – π and one

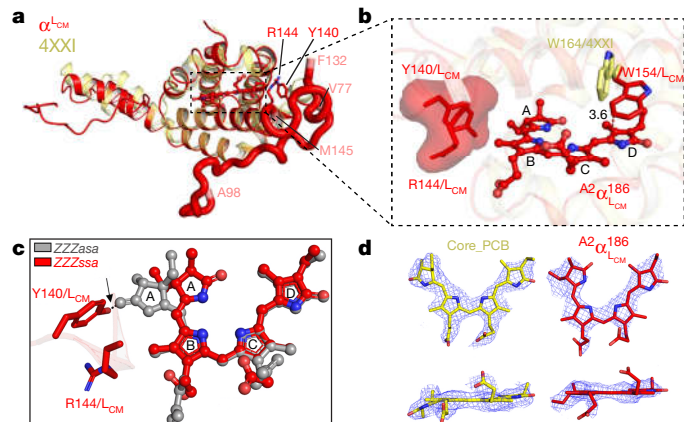


Fig. 4 | The conformation of PCB in L_{CM} . **a**, Structural alignment of the α subunit of L_{CM} ($\alpha^{L_{CM}}$) in the PBS and the recombinant α subunit of L_{CM} (PDB: 4XXI). $\alpha^{L_{CM}}$ and 4XXI are coloured red and pale yellow, respectively. The bilin in $\alpha^{L_{CM}}$ is shown in ball-and-stick representation. Two extra loops (V77–A98 and F132–M145) are displayed in sausage representation. **b**, Structural differences between $\alpha^{L_{CM}}$ and 4XXI. Two different conformations of tryptophan (W154 in L_{CM} and W164 in 4XXI) are shown in stick representation in red and yellow. Y140 and R144 from one loop are shown in surface representation in red. **c**, Steric clashes are observed between Y140 from L_{CM} and the ZZZssa configuration of $A^2\alpha_{L_{CM}}^{186}$ (grey), but is absent between Y140 and the ZZZssa configuration of $A^2\alpha_{L_{CM}}^{186}$ (red). **d**, Cryo-EM densities (mesh) of the bilins (stick) in the α subunit of the core (Core_PCB, yellow) and $\alpha^{L_{CM}}$ ($A^2\alpha_{L_{CM}}^{186}$, red) show the enhanced coplanarity of rings A and B of $A^2\alpha_{L_{CM}}^{186}$.

parallel-displaced π – π interaction with ring D, respectively. Moreover, R89, Y93 and Y97 interact with each other by either cation– π or π – π interactions (Fig. 3b). The superimposition of ApcF molecules from *P. purpureum* and *G. pacifica* with the β subunit from the core A2 shows that these three residues exist in the same position in all proteins, suggesting their importance for the stability of the PCB (Extended Data Fig. 8b). Except for these common features, an aromatic residue from ApcF (F60) is located above ring A of $A^2\beta_{ApcF}^{87}$ (Fig. 3b); an aromatic residue was also found in the ApcF of *G. pacifica* (Y60), whereas this residue is replaced by L60 in other β subunits (Extended Data Fig. 8b). This aromatic residue may therefore form additional π – π interactions with $A^2\beta_{ApcF}^{87}$ and hence lower its energy. Another notable feature in our structure of the complete PBS is that L_{CM} is directly involved in the interaction with $A^2\beta_{ApcF}^{87}$. Several hydrophobic residues of L_{CM} are located at the ApcF/ L_{CM} interface and within 4 Å of $A^2\beta_{ApcF}^{87}$ —this creates a contiguous hydrophobic ‘cap’ that buries $A^2\beta_{ApcF}^{87}$ (Fig. 3c, d, Extended Data Fig. 8c), which can enhance the stability of the conformation of $A^2\beta_{ApcF}^{87}$. The residues present around the chromophores from other β subunits ($A^2\beta_1^{81}$ and $A^2\beta_2^{81}$) are less hydrophobic (Extended Data Fig. 8d). Similarly, this hydrophobic cap is also found around the PCB of ApcF from the *G. pacifica* PBS (Extended Data Fig. 8e).

The terminal chromophore PCB in L_{CM} ($A^2\alpha_{L_{CM}}^{186}$) exhibits fluorescence with similar emission wavelengths to those of the intact PBS, and is at a lower energy than the upstream PCBs¹⁰. Although the overall structure of the α -subunit domain of L_{CM} ($\alpha^{L_{CM}}$) overlapped well with the recombinant $\alpha^{L_{CM}}$ (PDB: 4XXI)¹⁰ (Fig. 4a), some differences and new structural information are revealed in this study of the native PBS. In structure 4XXI, two different conformations of W164 are found above $A^2\alpha_{L_{CM}}^{186}$; one is parallel to ring D and the other is nearly perpendicular¹⁰. However, in our structure, W154 at the same position displays only one conformation, parallel to ring D (Fig. 4b)—this indicates that the native L_{CM} has a unique preference for how the side chain of such a residue is positioned. Moreover, compared with 4XXI, two extra loops (V77–A98 and F132–M145) were resolved in our $\alpha^{L_{CM}}$ (Figs. 1c, 5a). Two residues in this loop, Y140 and R144, are in direct contact (less than 4 Å) with $A^2\alpha_{L_{CM}}^{186}$.

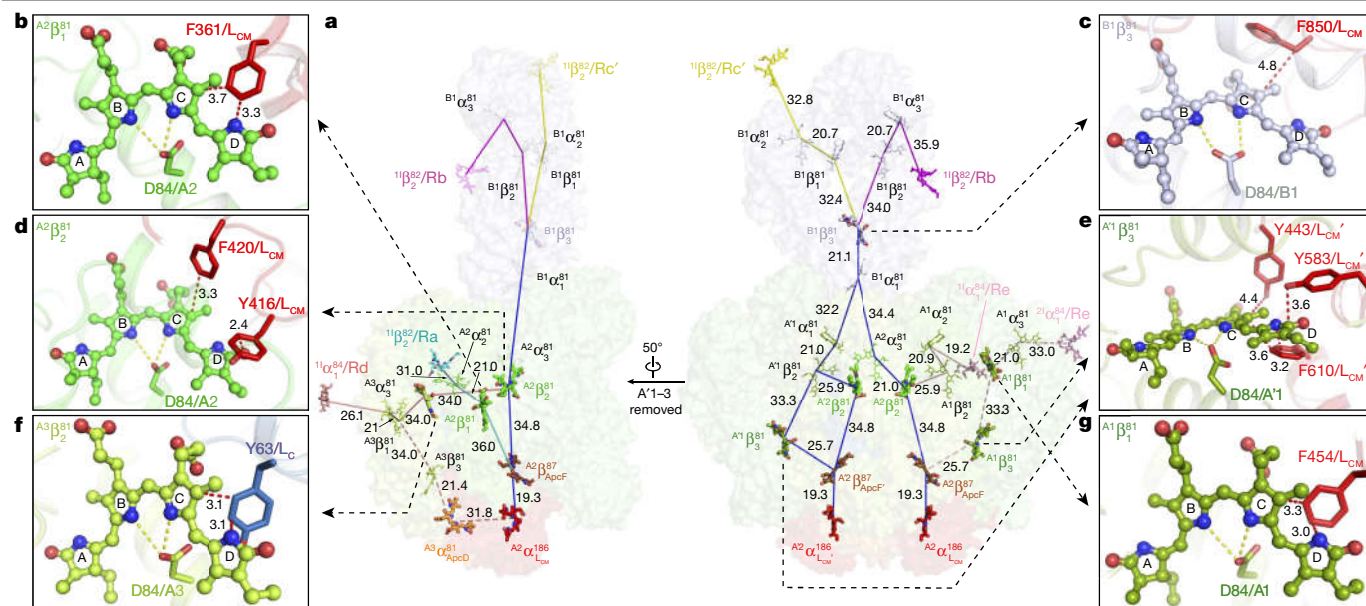


Fig. 5 | Key bilins in the core. **a**, Different views of the core with the bilins in different core layers shown in different colours. Bilins in $\alpha^{L_{CM}}$ ($\alpha^{L_{CM}}$), ApcD (ApcD') and ApcF (ApcF') are coloured red, orange and brown, respectively. Bilins in rods are the same colours as the rods in which they are located according to the colouring scheme in Fig. 1d. Bilins are shown in stick representation, and key bilins are shown as thicker sticks. The numbers

indicate the distances (Å) between the bilin pairs. **b**, Interaction between F361 from L_{CM} and the bilin $A2\beta_1^{81}$. **c**, Interaction between F850 from L_{CM} and the bilin $B1\beta_3^{81}$. **d**, Interactions between Y416 and F420 from L_{CM} with the bilin $A2\beta_1^{81}$. **e**, Interactions between Y443, Y583 and F610 from L_{CM}' with the bilin $A1\beta_3^{81}$. **f**, Interaction between Y63 from L_C and the bilin $A3\beta_2^{81}$. **g**, The interaction between F454 from L_{CM} and the bilin $A1\beta_1^{81}$.

(Fig. 4b). In particular, the side chain of Y140 is oriented towards the inside of the $A2\alpha_{L_{CM}}^{186}$ pocket (Fig. 4b). In such a conformation there will exist a steric clash between Y140 and the ZZZssa configuration of $A2\alpha_{L_{CM}}^{186}$, thus providing a driving force for the formation of the ZZZssa configuration (Fig. 4c, Extended Data Fig. 8f). Comparison of $\alpha^{L_{CM}}$ with five other similar subunits reveals that the orientation of this tyrosine in $\alpha^{L_{CM}}$ is opposite to that in other subunits (Extended Data Fig. 8g). Therefore, Y140 of L_{CM} is another factor that causes $A2\alpha_{L_{CM}}^{186}$ to uniquely adopt the ZZZssa geometry, which exhibits enhanced coplanarity of rings A and B compared with other PCBs in the α subunits of the core (Fig. 4d, Extended Data Fig. 3c).

In addition to $A3\alpha_{ApcD'}^{81}$, $A2\beta_{ApcF'}^{87}$ and $A2\alpha_{L_{CM}}^{186}$, the energy states of some other chromophores in the core are subjected to modification by the linker proteins. The shortest distance between rod Ra and the core was found between $11\beta_2^{82}/Ra$ and $A2\alpha_1^{81}$ (31 Å) (Fig. 5a), which may facilitate energy transfer from Ra to the core. The bilin nearest to $A2\alpha_1^{81}$ is $A2\beta_1^{81}$, rings C and D of which form a parallel-displaced π - π interaction with F361 from L_{CM} (Fig. 5b), and thus may mediate energy transfer to $A2\beta_{ApcF'}^{87}$. The energy absorbed by rods Rb and Rc' may travel through the core layer B1 to $B1\alpha_1^{81}$, the nearest bilin to the basal cylinders (Fig. 5a). Bilin $B1\beta_3^{81}$ may play an essential role in this process because it is subjected to a parallel-displaced π - π interaction with F850 from L_{CM} (Fig. 5c). In the basal cylinders, the two bilins $A2\beta_2^{81}/A2\beta_2^{81}$ and $A1\beta_3^{81}/A1\beta_3^{81}$ which are adjacent to the bilins on ApcF/ApcF' and separated from them by 34.8 Å and 25.7 Å, respectively—have special microenvironments (Fig. 5a). $A2\beta_2^{81}$ is affected by the π - π interactions with Y416 and F420 from L_{CM} (Fig. 5d) and $A1\beta_3^{81}$ is affected by several π - π interactions between its rings C and D with Y443, Y583 and F610 from L_{CM}' (Fig. 5e)—this suggests that these two bilins may facilitate energy flow to $A2\beta_{ApcF'}^{87}/A2\beta_{ApcF'}^{87}$. The bilin pair $Rd11\alpha_1^{84}$ and $A3\alpha_3^{81}$ shows the shortest distance (26.1 Å) between rod Rd and the core (Fig. 5a). The bilin $A3\beta_2^{81}$ may mediate further energy transfer because it is subjected to π - π interactions with Y63 from L_C (Fig. 5f). For Re, both hexamers Re1 and Re2 attach to the core layer A1; as such, the energy could flow from either Re1 or Re2 to the core. The shortest distance between Re2 and the core is from $21\alpha_1^{84}/Re$ to $A1\alpha_3^{81}$

(33 Å) (Fig. 5a). The energy could then travel via the $A1\beta_1^{81}$, because this bilin has the shortest distance to $A1\alpha_3^{81}$ and is affected by F454 from L_{CM} through two parallel-displaced π - π interactions with rings C and D (Fig. 5g). Together, our results show that core linker proteins are extensively involved in the modulation of the energy states of core bilins to ensure the efficient unidirectional transfer of energy. These findings provide the framework for a detailed examination of energy transfer in future studies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2020-7>.

- Bryant, D. A. & Canniffe, D. P. How nature designs light-harvesting antenna systems: design principles and functional realization in chlorophototrophic prokaryotes. *J. Phys. B* **51**, 033001 (2018).
- Adir, N., Dines, M., Klartag, M., McGregor, A. & Melamed-Frank, M. in *Complex Intracellular Structures in Prokaryotes* (ed. Shively, J. M.) 47–77 (Springer, 2006).
- Glazer, A. N. Light guides. Directional energy transfer in a photosynthetic antenna. *J. Biol. Chem.* **264**, 1–4 (1989).
- Adir, N., Bar-Zvi, S. & Harris, D. The amazing phycobilisome. *Biochim. Biophys. Acta Bioenerg.* <https://doi.org/10.1016/j.bbabi.2019.07.002> (2019).
- Zhang, J. et al. Structure of phycobilisome from the red alga *Griffithsia pacifica*. *Nature* **551**, 57–63 (2017).
- Sidler, W. A. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 139–216 (Springer, 1994).
- Singh, N. K., Sonani, R. R., Rastogi, R. P. & Madamwar, D. The phycobilisomes: an early requisite for efficient photosynthesis in cyanobacteria. *EXCLI J.* **14**, 268–289 (2015).
- Gao, X. et al. Molecular insights into the terminal energy acceptor in cyanobacterial phycobilisome. *Mol. Microbiol.* **85**, 907–915 (2012).
- Lundell, D. J., Yamanaka, G. & Glazer, A. N. A terminal energy acceptor of the phycobilisome: the 75,000-dalton polypeptide of *Synechococcus* 6301 phycobilisomes—a new biliprotein. *J. Cell Biol.* **91**, 315–319 (1981).
- Tang, K. et al. The terminal phycobilisome emitter, L_{CM} : a light-harvesting pigment with a phytochrome chromophore. *Proc. Natl Acad. Sci. USA* **112**, 15880–15885 (2015).

11. Glazer, A. N. & Bryant, D. A. Allophycocyanin B (λ_{max} 671, 618 nm): a new cyanobacterial phycobiliprotein. *Arch. Microbiol.* **104**, 15–22 (1975).
12. Peng, P. P. et al. The structure of allophycocyanin B from *Synechocystis* PCC 6803 reveals the structural basis for the extreme redshift of the terminal emitter in phycobilisomes. *Acta Crystallogr. D* **70**, 2558–2569 (2014).
13. Dong, C. et al. ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium *Synechococcus* sp. PCC 7002. *Biochim. Biophys. Acta* **1787**, 1122–1128 (2009).
14. Ashby, M. K. & Mullineaux, C. W. The role of ApcD and ApcF in energy transfer from phycobilisomes to PS I and PS II in a cyanobacterium. *Photosynth. Res.* **61**, 169–179 (1999).
15. Bryant, D. A., Guglielmi, G., de Marsac, N. T., Castets, A.-M. & Cohen-Bazire, G. The structure of cyanobacterial phycobilisomes: a model. *Arch. Microbiol.* **123**, 113–127 (1979).
16. Williams, R. C., Gingrich, J. C. & Glazer, A. N. Cyanobacterial phycobilisomes. Particles from *Synechocystis* 6701 and two pigment mutants. *J. Cell Biol.* **85**, 558–566 (1980).
17. Yamanaka, G., Glazer, A. N. & Williams, R. C. Molecular architecture of a light-harvesting antenna. Comparison of wild type and mutant *Synechococcus* 6301 phycobilisomes. *J. Biol. Chem.* **255**, 11104–11110 (1980).
18. Ducret, A., Sidler, W., Wehrli, E., Frank, G. & Zuber, H. Isolation, characterization and electron microscopy analysis of a hemidisoidal phycobilisome type from the cyanobacterium *Anabaena* sp. PCC 7120. *Eur. J. Biochem.* **236**, 1010–1024 (1996).
19. Arteni, A. A., Ajlani, G. & Boekema, E. J. Structural organisation of phycobilisomes from *Synechocystis* sp. strain PCC6803 and their interaction with the membrane. *Biochim. Biophys. Acta* **1787**, 272–279 (2009).
20. Chang, L. et al. Structural organization of an intact phycobilisome and its association with photosystem II. *Cell Res.* **25**, 726–737 (2015).
21. Gantt, E. & Lipschultz, C. A. Phycobilisomes of *Porphyridium cruentum*. I. Isolation. *J. Cell Biol.* **54**, 313–324 (1972).
22. Arteni, A. A. et al. Structure and organization of phycobilisomes on membranes of the red alga *Porphyridium cruentum*. *Photosynth. Res.* **95**, 169–174 (2008).
23. Gantt, E. & Lipschultz, C. A. Structure and phycobiliprotein composition of phycobilisomes from *Griffithsia pacifica* (Rhodophyceae). *J. Phycol.* **16**, 394–398 (1980).
24. Guglielmi, G., Cohen-Bazire, G. & Bryant, D. A. The structure of *Gloeobacter violaceus* and its phycobilisomes. *Arch. Microbiol.* **129**, 181–189 (1981).
25. Lange, W., Wilhelm, C., Wehrmeyer, W. & Mörschel, E. The supramolecular structure of photosystem II–phycobilisome-complexes of *Porphyridium cruentum*. *Bot. Acta* **103**, 250–257 (1990).
26. Gantt, E. & Conti, S. F. The ultrastructure of *Porphyridium cruentum*. *J. Cell Biol.* **26**, 365–381 (1965).
27. Bhattacharya, D. et al. Genome of the red alga *Porphyridium purpureum*. *Nat. Commun.* **4**, 1941 (2013).
28. Tajima, N. et al. Analysis of the complete plastid genome of the unicellular red alga *Porphyridium purpureum*. *J. Plant Res.* **127**, 389–397 (2014).
29. Gantt, E. & Lipschultz, C. A. Phycobilisomes of *Porphyridium cruentum*: pigment analysis. *Biochemistry* **13**, 2960–2966 (1974).
30. Glazer, A. N. & Hixson, C. S. Subunit structure and chromophore composition of rhodophytan phycoerythrins. *Porphyridium cruentum* B-phycoerythrin and b-phycoerythrin. *J. Biol. Chem.* **252**, 32–42 (1977).
31. Redlinger, T. & Gantt, E. Phycobilisome structure of *Porphyridium cruentum*: polypeptide composition. *Plant Physiol.* **68**, 1375–1379 (1981).
32. Ficner, R. & Huber, R. Refined crystal structure of phycoerythrin from *Porphyridium cruentum* at 0.23-nm resolution and localization of the gamma subunit. *Eur. J. Biochem.* **218**, 103–106 (1993).
33. Ducret, A., Sidler, W., Frank, G. & Zuber, H. The complete amino acid sequence of R-phycoerythrin- α and β subunits from the red alga *Porphyridium cruentum*. Structural and phylogenetic relationships of the phycocyanins within the phycobiliprotein families. *Eur. J. Biochem.* **221**, 563–580 (1994).
34. Glazer, A. N. & Hixson, C. S. Characterization of R-phycoerythrin. Chromophore content of R-phycoerythrin and C-phycoerythrin. *J. Biol. Chem.* **250**, 5487–5495 (1975).
35. Camara-Artigas, A. et al. pH-dependent structural conformations of B-phycoerythrin from *Porphyridium cruentum*. *FEBS J.* **279**, 3680–3691 (2012).
36. Ritter, S., Hiller, R. G., Wrench, P. M., Welte, W. & Diederichs, K. Crystal structure of a phycocouobilin-containing phycoerythrin at 1.90-Å resolution. *J. Struct. Biol.* **126**, 86–97 (1999).
37. Lüning, K. in *Seaweeds: Their Environment, Biogeography, and Ecophysiology*. (ed. Yarish, C. and Kirkman, H.) (John Wiley & Sons, 1990).
38. Liu, L.-N., Chen, X.-L., Zhang, Y.-Z. & Zhou, B.-C. Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red algae: an overview. *Biochim. Biophys. Acta* **1708**, 133–142 (2005).
39. Jiang, T., Zhang, J. & Liang, D. Structure and function of chromophores in R-phycoerythrin at 1.9 Å resolution. *Proteins* **34**, 224–231 (1999).
40. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry: Part II Techniques for the Study of Biological Structure and Function* (W. H. Freeman, 1980).
41. Gervasio, F. L., Chelli, R., Marchi, M., Procacci, P. & Schettino, V. Determination of the potential of mean force of aromatic amino acid complexes in various solvents using molecular dynamics simulations: the case of the tryptophan–histidine pair. *J. Phys. Chem. B* **105**, 7835–7846 (2001).
42. Gallivan, J. P. & Dougherty, D. A. Cation– π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
43. Gindt, Y. M., Zhou, J., Bryant, D. A. & Sauer, K. Spectroscopic studies of phycobilisome subcore preparations lacking key core chromophores: assignment of excited state energies to the L_{cm} , β^{B} and $\alpha^{\text{AP-B}}$ chromophores. *Biochim. Biophys. Acta* **1186**, 153–162 (1994).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Preparation of phycobilisomes

P. purpureum (From UTEX Culture Collection of Algae, UTEX2757) was cultured in Bold 1NV: Erdshreiber (1:1) half-seawater medium, bubbled with sterilizing filtered air at 22 °C, under a 16 h:8 h light–dark cycle, with a white-light flux of about 37 $\mu\text{mol photons per m}^2$ per second. Algal cells were collected by centrifugation for 10 min at 6,000g, and resuspended in Buffer A (0.65 M Na/KPO₄ buffer with 0.5 M sucrose and 10 mM EDTA, pH 7.0) at 0.3 g of wet weight per ml. Then cells were homogenized twice at 4 °C using a French Press (EmulsiFlex-C3, Avestin) at 4,000 p.s.i., and phenylmethylsulfonyl fluoride was added to a final concentration of 1 mM. After 30 min of incubation with lauryldimethylamine *N*-oxide (Sigma) (48 mg g⁻¹ wet algal cells), debris and supernatant chlorophyll were removed by centrifugation at 20,000g for 30 min at 18 °C. The middle aqueous violet solution was loaded in a discontinuous sucrose gradient (2 ml of 0.5 M, 2 ml of 0.75 M, 2 ml of 1.0 M, 2 ml of 1.5 M, 1 ml of 2.0 M, all in Buffer B: 0.75 M K/NaPO₄ buffer with 10 mM EDTA, pH 7.0) and spun at 120,000g for 4 h at 18 °C using a SW41 rotor on Optima XPN-100 centrifuge (Beckman Coulter). Three visible bands were obtained after centrifugation and violet band 1 is the main layer of intact PBSs (Extended Data Fig. 1a).

Absorption and fluorescence spectrum measurement

Absorption of the intact PBS was measured between 300–800 nm using an Ultrospec 2100 Pro ultraviolet–visible spectrophotometer (Biochrom).

Fluorescence emission spectra were recorded using a Hitachi FL-4500 fluorescence spectrophotometer at room temperature. After exciting at 450 nm, fluorescence emission was monitored from 500 to 700 nm.

Mass spectrometry analysis

Mass spectrometry analysis was performed as previously described⁵. In brief, the purified PBS was separated by 4–12% Bis-Tris SDS–PAGE in MES buffer and the gel was stained with ZnSO₄ to detect bilin-containing proteins with ultraviolet light by Zn-enhanced fluorescence. Then, four fluorescence bands with molecular mass greater than 25 kDa were excised for in-gel digestion and proteins were identified by mass spectrometry (Extended Data Fig. 1b). The intact PBS complex in solution was also subjected to mass spectrometric analysis. Finally, all 25 protein components of PBS were identified in the samples (Extended Data Fig. 1e).

Cryo-EM sample preparation and data collection

We used holey-carbon copper grids (Quantifoil R2/2, 400 mesh) covered with homemade ultrathin carbon for cryo-EM sample preparation. Cryo-EM grids were prepared with Vitrobot Mark IV (FEI Company) at 16 °C and 100% humidity. The grids were glow-discharged after adding 50 μl amylamine to a glass culture dish in the plasma cleaner and vapouring into the air. We added a 1.5 μl aliquot of protein with a concentration of 1.5 mg ml⁻¹ to the grids and waited for 60 s, and then added 3.5 μl of 50 mM Tris buffer (pH 8.0) to the grids and quickly mixed with the sample twice to reduce the salt concentration. The grids were then blotted for 3.5 s and plunged into liquid ethane cooled by liquid nitrogen.

The cryo-EM data were collected using a Titan Krios Microscope (FEI) operated at a voltage of 300 kV and equipped with a Cs corrector, a GIF Quantum energy filter (Gatan) and a K2 Summit direct electron detector (Gatan). A preset defocus range of $-1.2 \mu\text{m}$ to $-2.2 \mu\text{m}$ was used. All cryo-EM images were recorded at a nominal magnification of 105,000 \times in super-resolution mode. Each stack was exposed for 5.6 s with an exposure time of 0.175 s per frame and recorded as a movie of

32 frames, resulting in the total dose rate of approximately 48 electrons per \AA^2 for each stack. GIF was set to a slit width of 20 eV. The data were collected automatically using the software AutoEMation⁴⁴. The stacks were motion-corrected with MotionCor2⁴⁵ and binned twofold, resulting in a pixel size of 1.091 \AA per pixel.

Cryo-EM data analysis

A total of 16,218 micrographs were collected. Micrograph screening, manual particle picking and normalization were performed using EMAN2⁴⁶ and RELION3.0 beta^{47–49}. The contrast transfer function parameters of each micrograph were estimated using CTFFIND4^{50,51} and automatic particle picking; all the 2D, 3D classification, 3D refinement and local defocus calculation were performed with RELION3.0 beta^{47–49}.

The workflow of the data analysis is shown in Extended Data Fig. 2f. Two batches of data were collected and processed individually at the beginning. Particles were first manually picked from a small set of micrographs to produce templates for autopicking. Then particles were autopicked on all micrographs and manually screened to eliminate aggregation and ice contamination. Finally 322,889 and 363,480 particles were selected for the next 2D classification. After several rounds of 2D classification, 299,888 and 333,012 particles were left for the 3D classification. After 3D classification, two classes from each dataset with good quality were selected for the final reconstruction. At this point, we calculated the local defocus values for each particle and re-extracted particles from the dose-weighted micrographs⁵². Then two batches of data were merged to perform the 3D refinement. The final resolution of the 3D auto-refinement after post-processing was 2.82 \AA with a final particle number of 191,825 after imposing the C2 symmetry. Application of a mask for the core region during refinement further improved the resolution of this region to 2.68 \AA . We also applied local masks for each rod and extra phycocerythrin hexamer, which resulted in improved quality of local maps with resolutions ranging between 2.77 \AA and 3.56 \AA . The maps for the target regions were extracted from the overall map by Chimera⁵³, and the masks were created by RELION3.0 beta^{47–49}. All resolutions were estimated with the gold-standard Fourier shell correlation 0.143 criterion with high-resolution noise substitution. All the local resolution maps were calculated using ResMap⁵⁴.

Model building and refinement

We searched the published genome and transcriptome database of *P. purpureum*²⁷ against the 25 protein sequences of the *G. pacifica* PBS using the Basic Local Alignment Search Tool (BLAST). A total of 24 homologues, including eight PBP proteins and 16 linker proteins were obtained by this procedure, and these proteins are used as the candidates for model building. Local maps generated by the different masks described above were used to facilitate the model building process.

Because the sequences of *P. purpureum* and *G. pacifica* have high homology with each other, we first docked the structures of the *G. pacifica* PBS proteins (PDB: 5Y6P)⁵ into the electron microscopy maps using Chimera⁵³. All the PBP proteins and most of the linker proteins were fitted well. Then the sequence assignments were guided by well-resolved bulky residues such as phenylalanine, tyrosine, tryptophan and arginine, and the sequences of the *G. pacifica* PBS were replaced with corresponding residues in the *P. purpureum* PBS in Coot⁵⁵ and every residue was examined and manually adjusted to better fit in the map. Some of the L_Rγ4 proteins could not be fitted well at the N-terminal region. We first built the C-terminal CBDγ domain as described above, and then performed the de novo building in Coot⁵⁵ with bulky residues as land markers as most of these residues were clearly visible in our cryo-EM maps. The linker protein located at the centre cavity of the hexamer Hd is L_Rγ6 in the *G. pacifica* PBS that contains the CBDγ domain; however, the density at this region in the *P. purpureum* PBS shows recognizable structural features of the Pfam00427 domain. Therefore, we named this linker protein L_R6 and first docked the structure of the Pfam00427 domain from L_R2 into the density. By carefully

examining the densities outside the Pfam00427 domain of L_R6, a YYW motif was unambiguously identified according to the clear side-chain densities. Then we obtained the full-length sequence of L_R6 by searching the published genome and transcriptome database of *P. purpureum* for the protein containing both the Pfam00427 domain and the YYW motif. The sequence of L_R2 was replaced with corresponding residues in L_R6 in Coot⁵⁵ and de novo atomic model building was conducted for the rest of the sequence in Coot⁵⁵. Finally, 25 protein sequences were identified and confirmed by good agreement of the side-chain information between the sequences and the density maps (Supplementary Table 1).

The initial model was completed via iterative rounds of manual building with Coot⁵⁵ and refinement with phenix.real_space_refine^{56,57}. During this process, each part of the whole PBS model corresponding to each local map was refined against the local map with secondary structure and geometry restraints to prevent overfitting. Then, all parts were merged into a whole PBS model and this overall model was refined again against the overall 2.8 Å map using phenix.real_space_refine^{56,57}. The atomic model was cross-validated according to previously described procedures⁵⁸. In brief, atoms in the final model were randomly shifted by up to 0.5 Å, and the new model was then refined against one of two half-maps generated during the final 3D reconstruction. FSC values were calculated between the map generated from the resulting model and the two half-maps, as well as the averaged map of two half-maps. We did not observe notable separation between FSCwork and FSCfree, indicating that our model was not over-refined (Extended Data Fig. 2e). The data collection, model refinement and validation statistics are presented in Extended Data Tables 1, 2b. The statistics of the geometries of the models were generated using MolProbity⁵⁹. All the figures were prepared in PyMOL (<http://pymol.org>) or Chimera⁵³. The sequence alignments were performed by ClustalX2⁶⁰ and created by ESPript⁶¹.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The atomic coordinates have been deposited in the Protein Data Bank with the accession code 6KGX. The electron microscopy maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-9976 for the overall map and EMD-9977 through to EMD-9988 for the 12 local maps. The raw electron microscopy images used to build the 3D structure are available from the corresponding authors upon request.

44. Lei, J. & Frank, J. Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* **150**, 69–80 (2005).
45. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
46. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
47. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
48. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
49. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, 163 (2018).
50. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
51. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
52. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
53. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
56. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
57. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
58. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
59. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
60. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
61. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

Acknowledgements We thank the staff at the Tsinghua University Branch of the National Protein Science Facility (Beijing) for technical support on the Cryo-EM and High-Performance Computation platforms; J. Wang for model validation; D. Liu for model building and structure refinement; H. Lin and X. Pan for discussion; and X. Li and H.-W. Wang for recommendations for the computation. This work was supported by the National Basic Research Program (grants 2016YFA0501101 and 2017YFA0504600 to S.-F.S.) and the National Natural Science Foundation of China (grants 31670745 and 31861143048 to S.-F.S.).

Author contributions S.-F.S. supervised the project; J.M. prepared the samples, collected and analysed the electron microscopy data, performed the initial model building and the biochemical and biophysical analyses; X.Y. performed the model building and the structure refinement; J.M., X.Y., S.S. and S.-F.S. analysed the structure; X.W. helped with the electron microscopy data collection and the biochemical and biophysical analyses; S.Q. contributed to the sample selection; J.M. and X.Y. wrote the initial draft; and S.S. and S.-F.S. edited the manuscript.

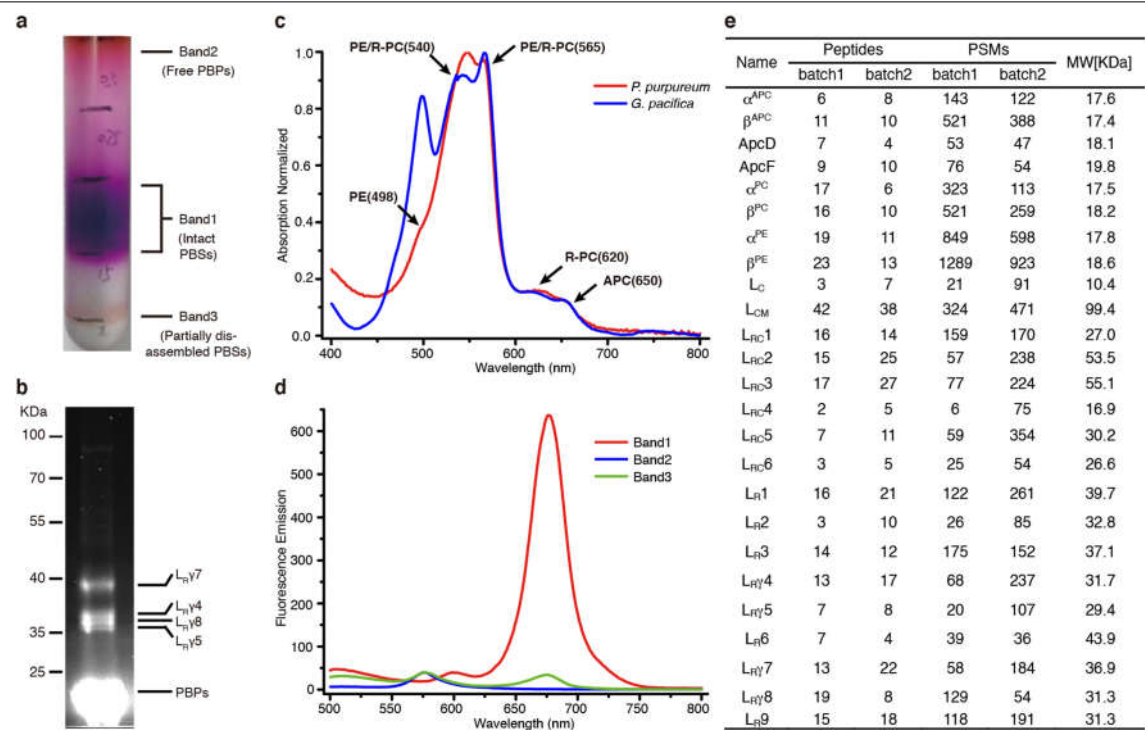
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2020-7>.

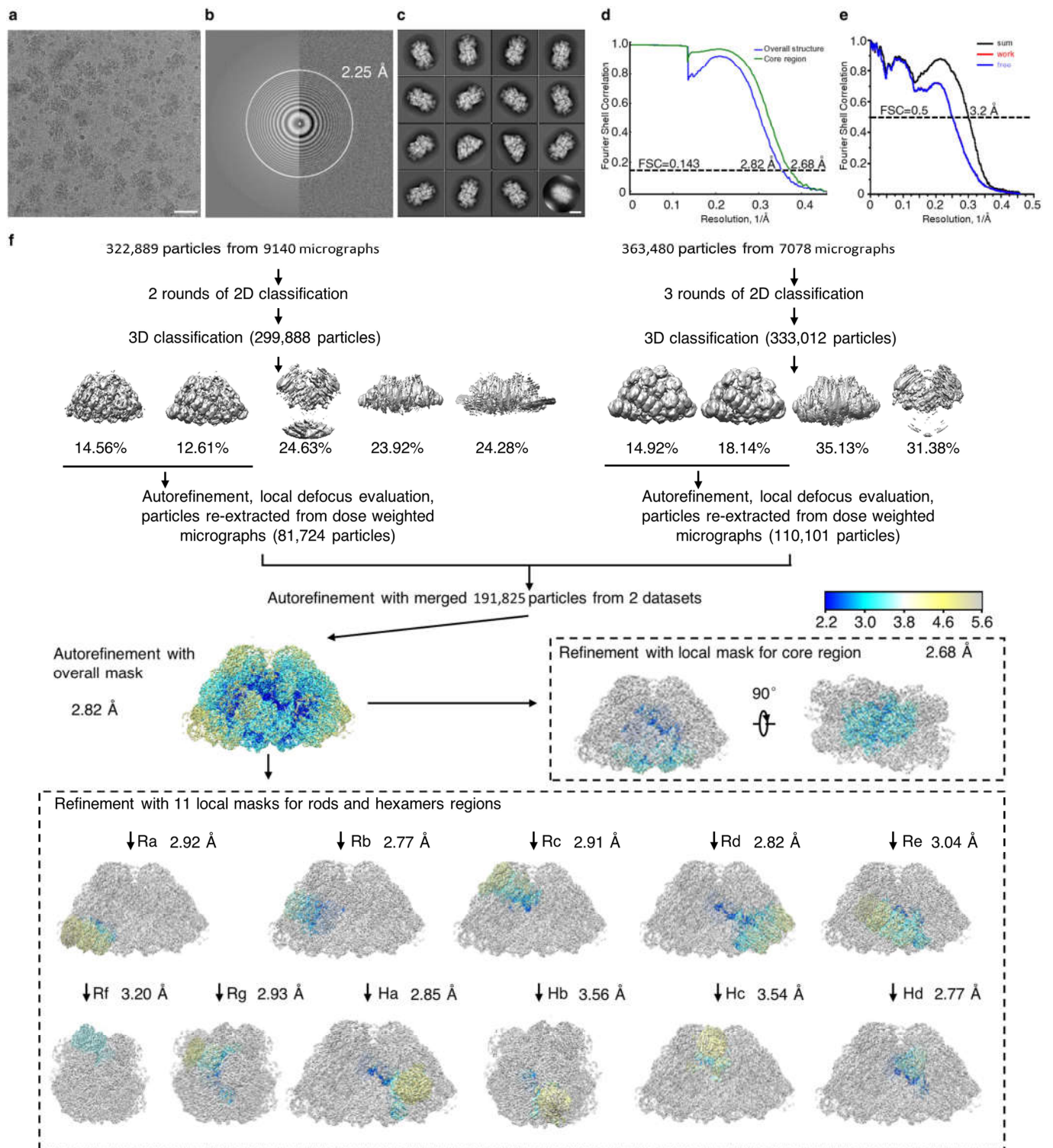
Correspondence and requests for materials should be addressed to S.-F.S.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



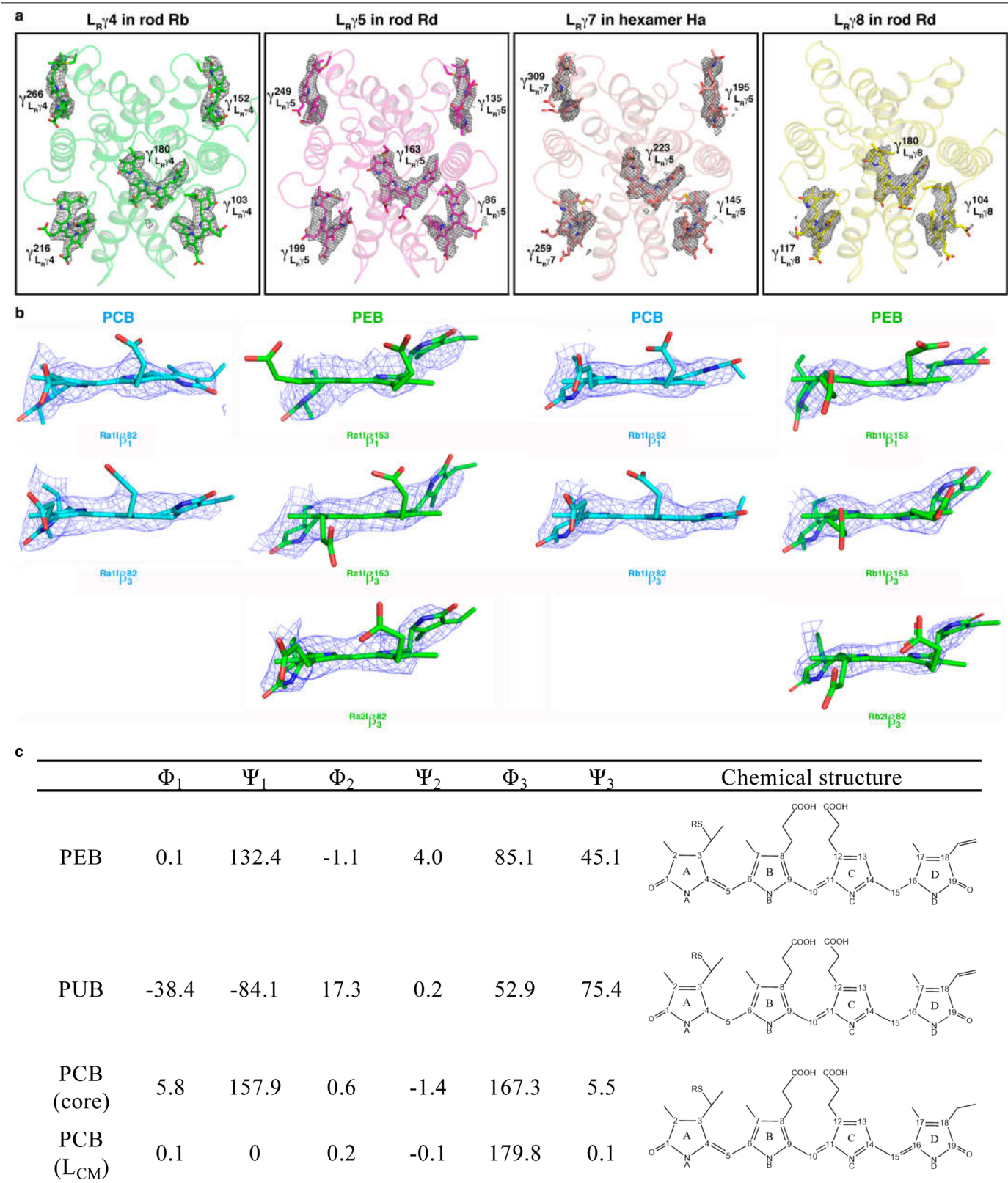
Extended Data Fig. 1 | Preparation and characterization of the PBS from *P. purpureum*. **a**, Isolation of PBSs using sucrose density gradient centrifugation. Three visible bands were observed. Band 1 is the sample of PBSs used for single-particle analysis in this study. The purification of PBS was repeated independently at least three times with similar results. **b**, Analysis of the protein composition of band 1 by SDS-PAGE stained with ZnSO₄ to enable the detection of bilin-containing proteins with ultraviolet light by Zn-enhanced fluorescence. The bands of L_RY4, 5, 7, 8 and PBPs identified by mass spectrometric analysis are indicated. For gel source data, see Supplementary Fig. 1. The purification and characterization of the protein composition was repeated independently at least three times with similar results. **c**, Absorption spectrum of band 1 and the PBS from *G. pacifica*. The peaks at 498 nm, 620 nm and 650 nm are from phycoerythrins, PCBs of phycocyanins and PCBs of

allophycocyanins, respectively. The peaks at 540 nm and 565 nm are from PEBs. The reduced absorption of the *P. purpureum* PBS compared with the *G. pacifica* PBS at 498 nm indicates that the phycoerythrin content of *P. purpureum* is much lower than that of *G. pacifica*. **d**, Fluorescence emission spectra of the three bands. Emission maxima at 580 nm and 676 nm represent the disassembled phycoerythrin hexamer and the terminal emitter in the intact PBS, respectively. Band 1 has an emission peak at 676 nm, band 2 at 580 nm and band 3 has two emission peaks at 676 nm and 580 nm, indicating that band 1 contains intact PBSs, band 2 contains free PBPs and band 3 contains partially disassembled PBSs. **e**, Results of the mass spectrometric analysis of purified PBSs. Two batches of sample were analysed. The similar results confirmed the consistency of our purification method.



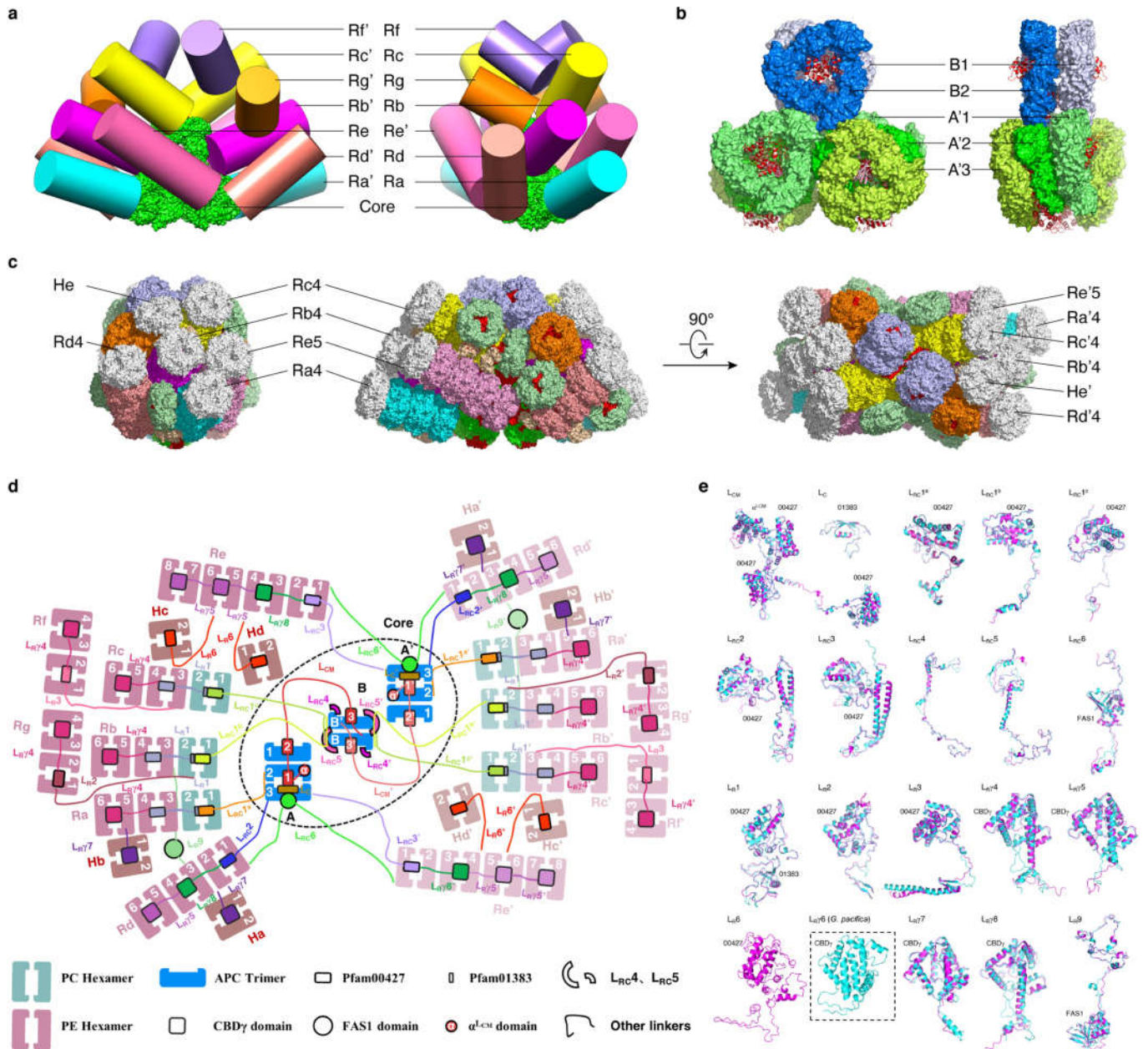
Extended Data Fig. 2 | Cryo-EM analysis of the PBS from *P. purpureum*. **a**, A representative motion-corrected electron micrograph of PBSs. Scale bar, 50 nm. Tens of thousands of micrographs were collected with similar results. **b**, Fourier power spectrum of the micrograph showing the Thon ring extending to 2.25 Å. Tens of thousands of micrographs were collected with similar results. **c**, Typical good, reference-free 2D class averages from single-particle PBS images. Scale bar, 20 nm. More than three rounds of 2D class average were performed with similar results. **d**, Gold-standard Fourier shell correlation (FSC)

curves for the 3D electron microscopy reconstructions of the PBS. Blue curve, FSC curve for the overall structure; green curve, FSC curve for the core region that was masked during refinement. **e**, FSC curves for the cross-validation of the atomic model. The small difference between work and free FSC curves suggested that the model was not overfitted. **f**, The workflow for the 2D and 3D classifications for cryo-EM data processing. The masking strategy for dealing with sub-regions of PBS is enclosed within dashed lines. For details, see 'Cryo-EM data analysis' in Methods.



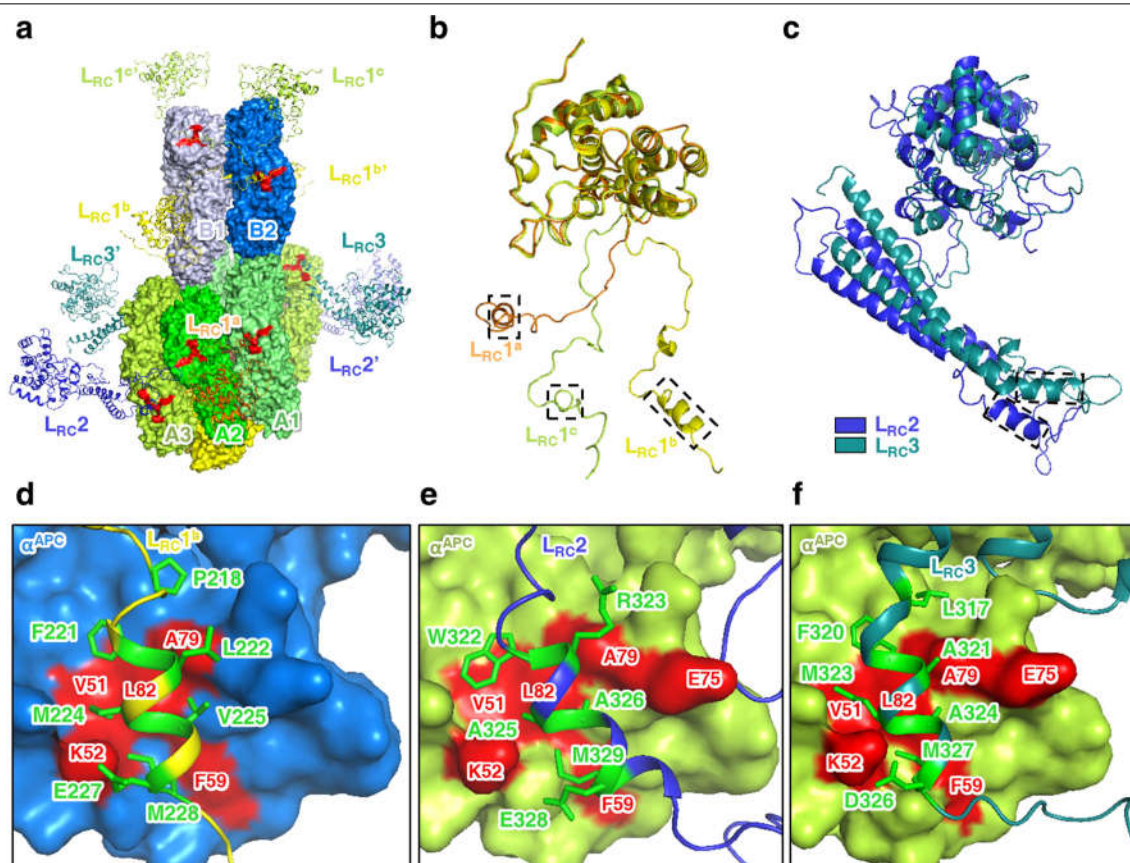
Extended Data Fig. 3 | Characterization of different types of chromophore. **a**, Cryo-EM densities (mesh) of bilins (stick) bound to $L_R\gamma 4$ in the rod Rb, $L_R\gamma 5$ in the rod Rd, $L_R\gamma 7$ in the hexamer Ha and $L_R\gamma 8$ in the rod Rd. **b**, The densities (mesh) of some PCB and PEB bilins (stick) in R-phycoyanins and phycoerythrins from rods Ra and Rb to show their different coplanarities. All of the density maps of PCB bilins showed a very flat conformation of rings B, C and D, consistent with the carbon-carbon double bond between rings C and D in PCB that constraints the movement of ring D, so that ring D is coplanar with the B-C plane.

Conversely, most of the density maps of PEB displayed a curved conformation of rings B, C and D owing to the single carbon-carbon bond between rings C and D in PEB that allows the rotation of ring D, so that ring D deviates from the B-C plane. However, some PEBs in R-phycoyanin also showed a planar conformation—such as $Ra11\beta_{153}$ and $Rb11\beta_{153}$ —although to a lesser extent than that for a typical PCB molecule. **c**, Dihedral angles of three kinds of chromophore. The dihedral angles $\Phi_1, \Psi_1, \Phi_2, \dots$ are defined by the atoms NA-C(4)-C(5)-C(6), C(4)-C(5)-C(6)-NB, NB-C(9)-C(10)-C(11) ... etc.



Extended Data Fig. 4 | Overall structure of the PBS from *P. purpureum* and comparison with that from *G. pacifica*. **a**, Schematic diagram showing the organization of the rods and the core from two perpendicular views. The colouring scheme is the same as in Fig. 1e. **b**, Structure of the core from two perpendicular views shows the assembly and arrangement of the core layers. **c**, Overall structure of the PBS overlapped with the *G. pacifica* PBS displayed in surface representation from three perpendicular views. The additional hexamers in the *G. pacifica* PBS are coloured white and labelled. **d**, Schematic model of the PBS architecture. The connections between PBS components are shown. Dark and light colours show C2 symmetric parts of rods. Dark and light salmon, phycoerythrin hexamers in rod; dark and light brown, extra phycoerythrin hexamers; dark and light forest green, phycocyanin hexamers; blue, allophycocyanin trimer; large rectangular box, Pfam00427 domains;

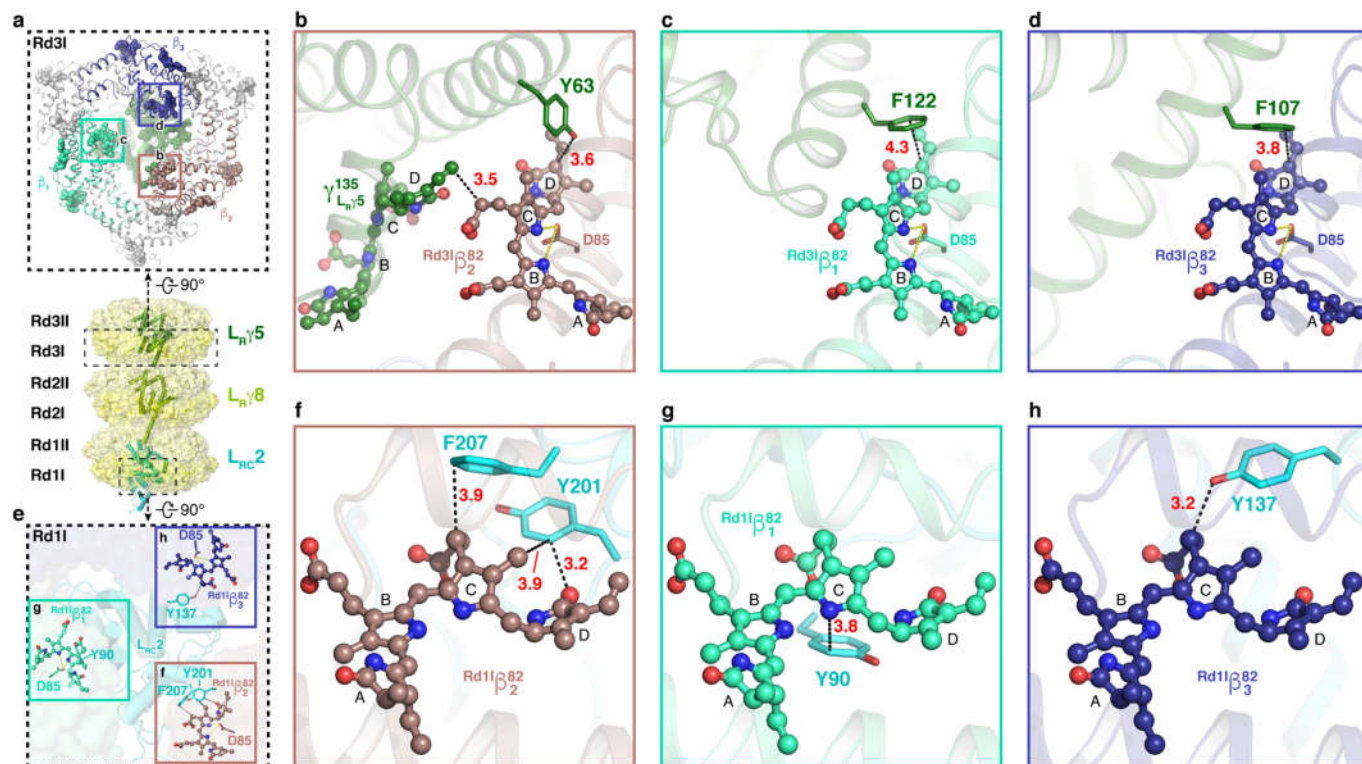
small rectangular box, Pfam01383 domains; square box, CBDγ. **e**, Comparison of linker proteins from *P. purpureum* with those from *G. pacifica*. Structures of the 19 well-resolved linker proteins (magenta) are superimposed with those from the *G. pacifica* PBS (cyan). The linker proteins share very high structural conservation—such as the Pfam00427 domain in the rod-core linker ($L_{RC}1-3/L_{RC}1'-3'$), the rod linker ($L_R1-3/L_R1'-3'$ and L_{CM}/L_{CM}'), the Pfam01383 domain in the core linker (L_C/L_C' and L_R1/L_R1'), the FAS1 domain in $L_{RC}6/L_{RC}6'$ and L_R9/L_R9' , the CBDγ domain in $L_{R\gamma}4-5/L_{R\gamma}4'-5'$ and $L_{R\gamma}7-8/L_{R\gamma}7'-8'$, the coiled-coil motif at the C termini of $L_{RC}2-3/L_{RC}2'-3'$, and the long α -helix in the middle of the $L_{RC}4-5/L_{RC}4'-5'$. Note that L_R6 from the *P. purpureum* PBS is different from L_R6 from the *G. pacifica* PBS, therefore they are not aligned. Domains of α^{1-CT} , Pfam00427 (00427), Pfam01383 (01383), CBDγ, and FAS1 are labelled.



Extended Data Fig. 5 | Interactions between L_{RC} proteins and the core.

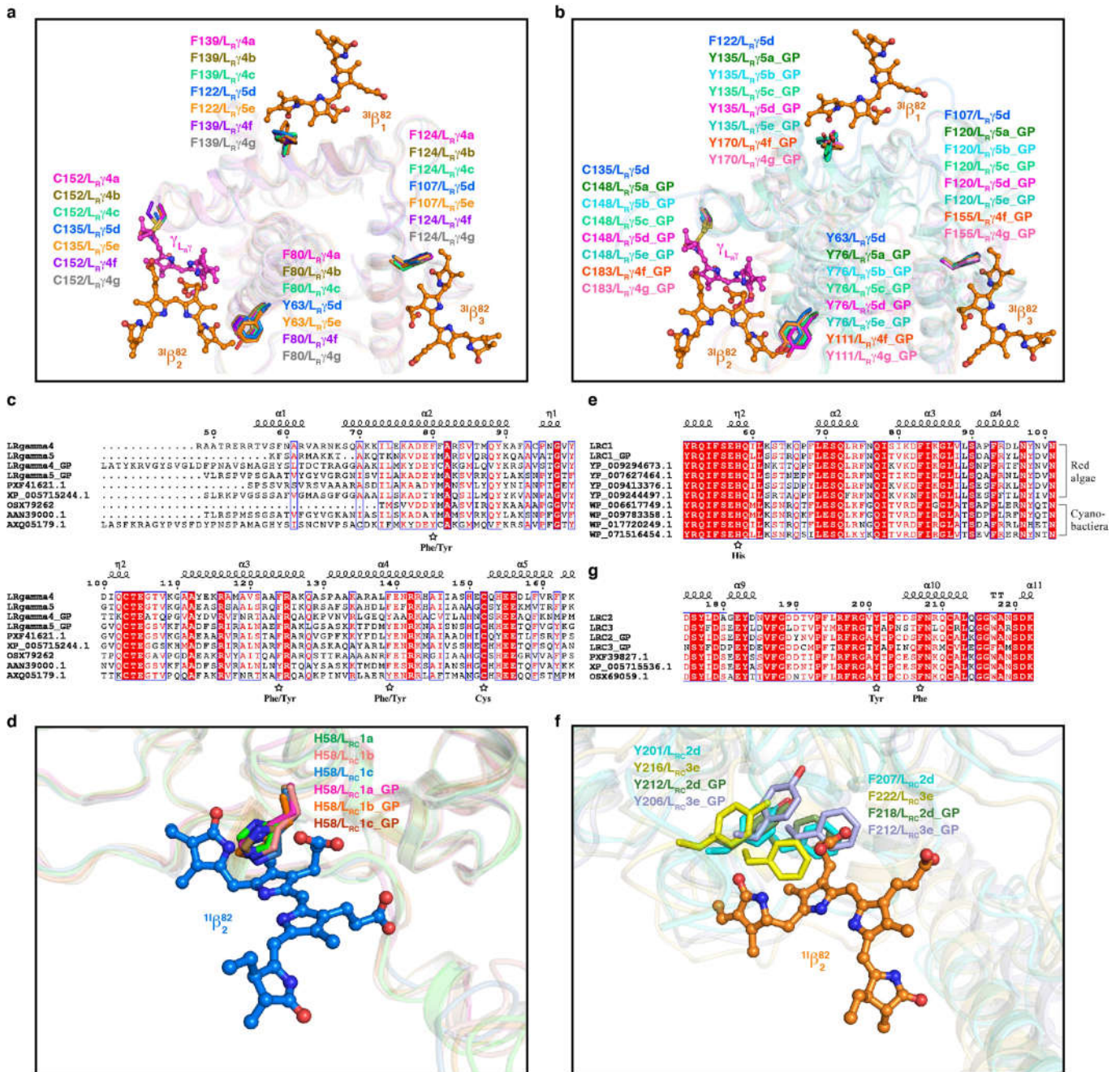
a, Organization of L_{RC} proteins L_{RC}1–3/L_{RC}1′–3′ and the core. The grooves on the α subunits that contact the linker helices are shown in red. **b**, Structural similarity and differences among L_{RC}1^a, L_{RC}1^b and L_{RC}1^c. These rod–core linkers are superimposed relative to the Pfam00427 domain. The helices that interact with the core are boxed. **c**, Structural similarity of L_{RC}2 and L_{RC}3, as

demonstrated by superimposition of the Pfam00427 domain at the N termini and the coiled-coil motif at the C termini. The helices interacting with the core are boxed. **d–f**, Interactions between the α^{APC} subunit and the helices of L_{RC}1^b (**d**), L_{RC}2 (**e**) and L_{RC}3 (**f**). The residues involved in the interaction of L_{RC} proteins are coloured green and shown in stick representation. The α^{APC} are shown in surface representation, and the residues involved in the interaction are red.



Extended Data Fig. 6 | Interactions of the linker proteins L_RYs and L_{RC}s with chromophores in the rod Rd. **a**, Bottom, overall structure of the rod Rd with the hexamers shown in surface representation and the linker proteins shown in cartoon representation. Top, structure of the layer Rd3I. Proteins and bilins are shown in cartoon and sphere representations, respectively. Three β subunits are coloured differently and the $\beta 82$ PEBs are boxed and analysed in detail in **b-d**. **b**, The interactions between the residue Y63 and the bilin Y_{LRY5}^{135} with the bilin $Rd3I\beta_2^{82}$. **c**, The interaction between F122 from L_RY5 and the bilin

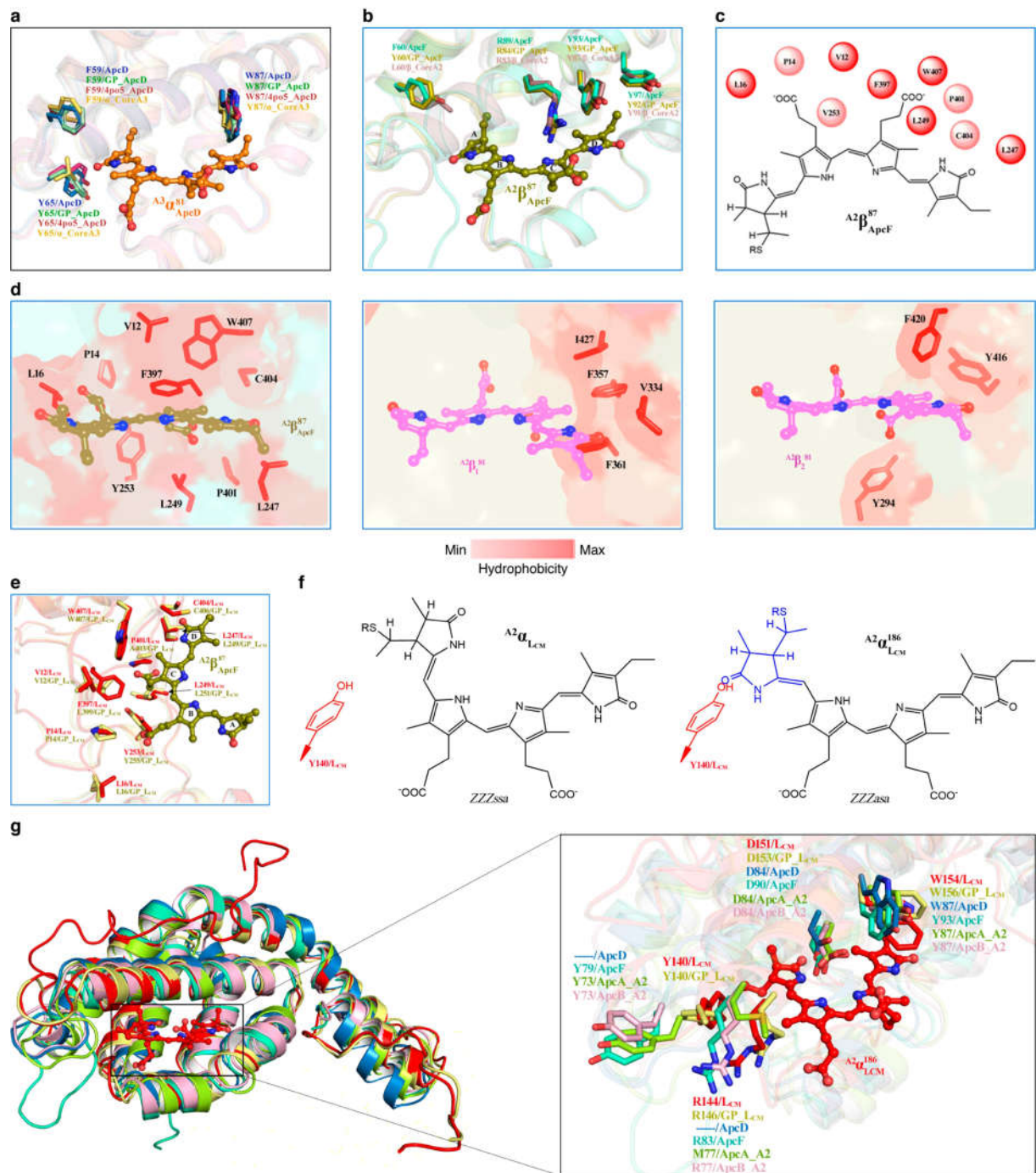
$Rd3I\beta_1^{82}$. **d**, The interaction between F107 from L_RY5 and the bilin $Rd3I\beta_3^{82}$. **e**, A focused view of the structure of the layer Rd1I showing the central triangle area. PBPs, the linker protein, bilins and residues are shown in surface, cartoon, ball-and-stick and stick representations, respectively. Three $\beta 82$ PCBs are boxed and analysed in detail in **f-h**. **f**, The interactions between Y201 and F207 from L_{RC}2 and the bilin $Rd1I\beta_2^{82}$. **g**, The interaction between Y90 from L_{RC}2 and the bilin $Rd1I\beta_1^{82}$. **h**, The interaction between Y137 from L_{RC}2 and the bilin $Rd1I\beta_3^{82}$.



Extended Data Fig. 7 | Comparisons of linker proteins from both

***P. purpureum* and *G. pacifica*.** **a, b**, Structural alignment of L_γ linker proteins in the outmost hexamers of various rods from the *P. purpureum* PBS (**a**) and the *G. pacifica* PBS (**b**). β82 PEBs and residues of L_γ linker proteins are shown in ball-and-stick and stick representations, respectively. Note that an aromatic residue from the L_γ linker is present near each β82 PEB to form π-π interactions, and one bilin from the L_γ linker (Y₆₃) always provides additional π electrons to the conjugation system of the β82 PEB. These aromatic residues and the bilins from L_γ linker proteins are conserved in both *P. purpureum* and *G. pacifica*. **c**, Sequence alignment of L_γ4-5 from *P. purpureum* and other red algae. Three aromatic residues interacting with the β82 PEBs and the cysteine residues used to link the bilins close to the β82 PEBs are marked by stars. LRgamma4_GP and LRgamma5_GP, L_γ4-5 from *G. pacifica*; PXF41621.1, γ-subunit from *Gracilariaopsis chorda*; XP_005715244.1, γ-subunit from *Chondrus crispus*; OSX79262, γ-subunit from *Porphyra umbilicalis*; AAN39000.1, γ-subunit from *Griffithsia japonica*; AXQ05179.1, γ-subunit from *Agarophyton chilense*. **d**, Structural alignment of L_γ4 linker proteins from *P. purpureum* and *G. pacifica* in the phycocyanin hexamer showing the bilin β82

and the surroundings. The key histidine residue close to the β82 PCB is conserved. **e**, Sequence alignment of L_{RC}1 from *P. purpureum* and other red algal and cyanobacterial species. The key histidine residue close to the β82 PCBs is marked with a star. LRC1_GP, L_{RC}1 from *G. pacifica*; YP_009294673.1, L_{RC}1 from red algal *G. chorda*; YP_007627464.1, L_{RC}1 from red algal *C. crispus*; YP_009413376.1, L_{RC}1 from red algal *P. umbilicalis*; YP_009244497.1, L_{RC}1 from red algal *A. chilense*; WP_006617749.1, L_{RC}1 from cyanobacteria *Arthrospira platensis*; WP_009783358.1, L_{RC}1 from cyanobacteria *Lyngbya* sp. PCC 8106; WP_017720249.1, L_{RC}1 from cyanobacteria *Oscillatoria* sp. PCC 10802; WP_071516454.1, L_{RC}1 from cyanobacteria *Geitlerinema* sp. PCC 9228. **f**, Structural alignment of the L_{RC}2 and L_{RC}3 linker proteins from *P. purpureum* and *G. pacifica* in the phycoerythrin hexamer proximal to the core showing the bilin β82 and the surroundings. Two aromatic residues near to the β82 PEB are conserved in both *P. purpureum* and *G. pacifica*. **g**, Sequence alignment of L_{RC}2-3 from *P. purpureum* and other red algae. Two aromatic residues close to the β82 PEBs are marked with stars. L_{RC}2_GP and L_{RC}3_GP are from *G. pacifica*. PXF39827.1, XP_005715536.1 and OSX69059.1 are from *G. chorda*, *C. crispus* and *P. umbilicalis*, respectively.



Extended Data Fig. 8 | Characterization of ApcD, ApcF and the α subunit domain of L_{CM} . **a**, Magnified view of the superimposition of ApcD proteins from *P. purpureum*, *G. pacifica* (GP_ApcD), *Synechocystis* PCC 6803 (4PO5_ApcD) and the α subunit of the core layer A3 (α _CoreA3). Bilins and residues are shown in ball-and-stick and stick representations, respectively. Three aromatic residues near the PCB are conserved in all ApcD proteins, but not in the α subunit of the core layer A3. **b**, Magnified view of the superimposition of ApcF proteins from *P. purpureum* and *G. pacifica* (GP_ApcF), and the β subunit of the core A2 (β _CoreA2). $A2\beta^{87}_{ApcF}$ is shown in ball-and-stick representation in sand. **c**, A schematic of interactions between $A2\beta^{87}_{ApcF}$ and the hydrophobic cap.

d, Magnified view of the PCB pocket of ApcF (left), $A2\beta^{81}_{ApcF}$ (middle) and $A2\beta^{81}_{ApcF}$ (right). The protein is shown in surface representation and coloured on the basis of amino acid hydrophobicity (see colour bar). The side chains of hydrophobic residues within 5 Å of the PCB are shown in stick representation. **e**, Magnified view of the structural alignment of the hydrophobic caps formed by L_{CM} proteins from *P. purpureum* and *G. pacifica*. **f**, Schematic of the steric hindrance experienced by Y140/ L_{CM} and the ZZZ_{ssa} configuration of $A2\alpha^{186}_{LCM}$. **g**, Structural alignment of α_{LCM} , ApcD, ApcF, the α subunit (ApcA_A2) and the β subunit (ApcB_A2) in the core. The PCB pockets are indicated in the magnified view on the right.

Article

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	Phycobilisome from <i>P. purpureum</i> (EMDB-9976) (PDB 6KGX)
Data collection and processing	
Magnification	105,000
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	48
Defocus range (μm)	-1.2 ~ -2.2
Pixel size (Å)	1.091
Symmetry imposed	C2
Initial particle images (no.)	686,369
Final particle images (no.)	191,825
Map resolution (Å)	2.82
FSC threshold	0.143
Map resolution range (Å)	2.4~7.4
Refinement	
Initial model used (PDB code)	5Y6P
Model resolution (Å)	2.82
FSC threshold	0.143
Model resolution range (Å)	2.4~7.4
Map sharpening <i>B</i> factor (Å ²)	-54.74
Model composition	
Non-hydrogen atoms	1014714
Protein residues	125577
Ligands	1598
<i>B</i> factors (Å ²)	
Protein	57.95
Ligand	71.06
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	2.064
Validation	
MolProbity score	1.79
Clashscore	8.07
Poor rotamers (%)	0.53
Ramachandran plot	
Favored (%)	94.88
Allowed (%)	5.11
Disallowed (%)	0.01

Extended Data Table 2 | Summary of proteins, chromophores and model validation

a

Subunit	Numbers in PBS	Chromophore						Total
		PCB		PEB		PUB		
		Per subunit	Total	Per subunit	Total	Per subunit	Total	
α^{APC}	20	1	20					20
β^{APC}	22	1	22					22
ApcD	2	1	2					2
ApcF	2	1	2					2
α^{PC}	36	1	36					36
β^{PC}	36	1	36	1	36			72
α^{PE}	254			2	508			508
β^{PE}	274			3	822			822
L _C	2							
L _{CM}	2	1	2					2
L _{RC1}	6							
L _{RC2}	2							
L _{RC3}	2							
L _{RC4}	2							
L _{RC5}	2							
L _{RC6}	2							
L _{R1}	6							
L _{R2}	2							
L _{R3}	2							
L _{RY4}	10			3	30	2	20	50
L _{RY5}	6			3	18	2	12	30
L _{R6}	4							
L _{RY7}	4			3	12	2	8	20
L _{RY8}	4			1	4	2	8	12
L _{R9}	2							
Total	706		120		1430		48	1598

b

Molecule*	MolProbability Scores	Ramachandran plot statistics (%)			RMS deviations	
		Favored	Allowed	Outliers	Bonds Length (Å)	Bonds Angles (°)
Core	1.52	96.53	3.47	0.00	0.004	1.518
Ra/Ra'	1.75	94.43	5.57	0.00	0.009	2.149
Rb/Rb'	1.39	96.32	3.68	0.01	0.007	1.919
Rc/Rc'	1.53	95.96	4.04	0.00	0.009	1.952
Rd/Rd'	1.64	95.77	4.23	0.00	0.012	2.076
Re/Re'	1.62	94.76	5.23	0.01	0.008	2.287
Rf/Rf'	1.78	93.33	6.67	0.00	0.007	1.919
Rg/Rg'	1.44	96.16	3.84	0.00	0.010	2.175
H/H'	1.65	94.64	5.31	0.05	0.013	2.247
M/M'	1.46	95.42	4.58	0.00	0.006	2.239

a. Numbers of proteins and chromophores in the PBS. **b.** Summary of model validation for the PBS components. *Core contains all α subunits, β subunits in core, and L_C1/L_C1' , L_{CM}/L_{CM}' , L_{RC4}/L_{RC4}' , L_{RC5}/L_{RC5}' and L_{RC6}/L_{RC6}' ; each rod (Ra/Ra'–Rg/Rg') contains all α subunits, β subunits and linker proteins in the rod; H/H' contains all extra hexamers (Ha/Ha', Hb/Hb', Hc/Hc' and Hd/Hd') including all α subunits, β subunits and linker proteins in the hexamers, and L_R9/L_R9' ; M/M' contains individual ($\alpha\beta$) monomers and all individual β subunits.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection AutoEMation 1.0

Data analysis Relion 2.0, Relion 3.0 beta, MotionCor2 1.1.0, EMAN2.1, CTFFIND4, Phenix 1.14-3260, Coot 0.8.9.1, Pymol 1.8.2.1, Chimera 1.12, CLUSTAL 2.0, ResMap v1.1, MolProbity 4.4, ENDscript 2.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The atomic coordinates have been deposited in the Protein Data Bank with the accession code 6KGX. The EM maps have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-9976 for the overall map and EMD-9977 to EMD-9988 for the twelve local maps. All other data and materials are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Amount of cryo-EM micrographs collected was based on the previous knowledge that the reconstruction of the protein particles picked from these micrographs could reach to a near atomic resolution and also limited by the time allocation of the microscope.
Data exclusions	The exclusion criteria were not pre-established. 2D and 3D classification yielded multiple classes. Only the particles in the classes that showed clear structural signals and intact structures were selected, combined and used in the final reconstruction and refinement. Details are described in the flowchart of Extended Data Figure 2e and Methods.
Replication	Multiple rounds of structural refinement have been performed and all resulted in same density maps (with different resolutions though). The purification and characterization of PBS (sucrose density gradient centrifugation, SDS-PAGE, absorption spectrum, fluorescence emission spectra) have been repeated for at least three times with similar results. Two batches of sample were analyzed by MS and the similar results confirmed the consistency of our purification method.
Randomization	Randomization of samples is not relevant for a single particle electron microscopy study because the study focused on a specific protein complex.
Blinding	Binding is not relevant because we are studying a specific protein complex.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging


Structural basis of ligand recognition and self-activation of orphan GPR52

<https://doi.org/10.1038/s41586-020-2019-0>

Received: 4 September 2019

Accepted: 23 January 2020

Published online: 19 February 2020

 Check for updates

Xi Lin^{1,2,3,4,13}, Mingyue Li^{1,2,3,4,13}, Niandong Wang^{1,2,3,4,13}, Yiran Wu^{1,13}, Zhipu Luo⁵, Shimeng Guo⁶, Gye-Won Han⁷, Shaobai Li^{8,9}, Yang Yue¹, Xiaohu Wei^{2,3,4}, Xin Xie^{2,6}, Yong Chen^{2,3,4}, Suwen Zhao^{1,2}, Jian Wu^{8,9,10}, Ming Lei^{8,9,11} & Fei Xu^{1,2,12}✉

GPR52 is a class-A orphan G-protein-coupled receptor that is highly expressed in the brain and represents a promising therapeutic target for the treatment of Huntington's disease and several psychiatric disorders^{1,2}. Pathological malfunction of GPR52 signalling occurs primarily through the heterotrimeric G_s protein², but it is unclear how GPR52 and G_s couple for signal transduction and whether a native ligand or other activating input is required. Here we present the high-resolution structures of human GPR52 in three states: a ligand-free state, a G_s-coupled self-activation state and a potential allosteric ligand-bound state. Together, our structures reveal that extracellular loop 2 occupies the orthosteric binding pocket and operates as a built-in agonist, conferring an intrinsically high level of basal activity to GPR52³. A fully active state is achieved when G_s is coupled to GPR52 in the absence of an external agonist. The receptor also features a side pocket for ligand binding. These insights into the structure and function of GPR52 could improve our understanding of other self-activated GPCRs, enable the identification of endogenous and tool ligands, and guide drug discovery efforts that target GPR52.

GPCRs are membrane proteins with 7 transmembrane helical domains, and over 800 members of this family are included in the human genome. Among them, more than 100 are orphan receptors—that is, receptors for which the endogenous ligands have not yet been identified⁴. GPR52 is a class-A orphan GPCR but exhibits a low sequence homology (less than 20%) to non-orphan GPCRs, which hinders an in-depth understanding of its structure and the discovery of any tool ligands. GPR52 has important roles in the brain, and is therefore an emerging target for the treatment of a variety of psychiatric diseases⁵. In particular, GPR52 colocalizes with the D₂ dopamine receptor (D2R) in the striatum of the basal ganglia and—through its G_s coupling activity—can antagonize D2R signalling by causing cellular accumulation of cAMP⁶. GPR52 agonists are therefore regarded as potential therapeutics for schizophrenia⁷, cognitive impairment⁸, psychiatric disorders^{2,9}, brain malformation⁵ and hyperactivity⁶. In addition, antagonists or inverse agonists of GPR52 are possible drug candidates for the treatment of Huntington's disease, as GPR52 is associated with the abnormal expression of huntingtin that is observed in patients with this disorder^{1,10}.

The physiological functions of GPR52 are all closely related to its coupling to the heterotrimeric G_s protein and the downstream signalling that results. Several structures of GPCRs in complex with G_s have previously been determined—including the β₂ adrenergic

receptor (β₂AR¹¹), adenosine A_{2A} receptor (A_{2A}R¹²), calcitonin receptor (CTR¹³), glucagon-like peptide 1 receptor (GLP-1R¹⁴), calcitonin gene-related peptide receptor (CGRP¹⁵) and parathyroid hormone receptor 1 (PTH1R¹⁶)—enabling the general landscape of G_s recognition by GPCRs to be elucidated across class-A and class-B GPCRs. All of these GPCRs require an agonist to induce a conformational change and form the agonist–GPCR–G-protein complexes. As GPR52 is an orphan GPCR, the identity of its native ligand remains unknown. However, a previous study showed that GPR52 when expressed on its own (that is, without any agonist) exhibited a high level of basal activity³—reminiscent of the 5-HT_{2C} serotonin receptor and several other understudied GPCRs that also show high levels of basal activity^{17,18}. It is unclear how this high level of intrinsic activity is achieved, and whether a native stimulator (that is, endogenous agonist) is required for G-protein coupling and signal transduction.

Here we present high-resolution structures of human GPR52 in the ligand-free and G_s-coupled states. The structures reveal that extracellular loop 2 (ECL2) of GPR52 occupies the orthosteric ligand-binding pocket and has an essential role in the self-activation of the receptor. We also report a structure of GPR52 in complex with a surrogate agonist, c17, and identify a ligand-binding pocket. Our results provide an integrated understanding of the structure and function of GPR52 and its mechanism of self-activation.

¹Human Institute, ShanghaiTech University, Shanghai, China. ²School of Life Science and Technology, ShanghaiTech University, Shanghai, China. ³Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. ⁴University of Chinese Academy of Sciences, Beijing, China. ⁵Institute of Molecular Enzymology, School of Biology and Basic Medical Sciences, Soochow University, Suzhou, China. ⁶CAS Key Laboratory of Receptor Research, National Center for Drug Screening, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China. ⁷Department of Biological Sciences, Bridge Institute, University of Southern California, Los Angeles, CA, USA. ⁸Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. ⁹Shanghai Institute of Precision Medicine, Shanghai, China. ¹⁰Shanghai Key Laboratory of Translational Medicine on Ear and Nose Diseases, Shanghai, China. ¹¹Key Laboratory of Cell Differentiation and Apoptosis, Chinese Ministry of Education, Shanghai Jiao Tong University School of Medicine, Shanghai, China. ¹²Centre for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai, China. ¹³These authors contributed equally: Xi Lin, Mingyue Li, Niandong Wang, Yiran Wu.

✉e-mail: wujian@shsmu.edu.cn; leim@shsmu.edu.cn; xufei@shanghaitech.edu.cn

Crystal structure of ligand-free GPR52

We first crystallized human GPR52 in the ligand-free (apo) state. The two apo GPR52 structures we present (GPR52-Rub-apo and GPR52-Fla-apo) were engineered with different intracellular loop 3 (ICL3) fusion-protein partners (rubredoxin (Rub) and flavodoxin (Fla)) and crystallized in different space groups. To further improve the stability and surface expression of GPR52, we introduced 7 point mutations to the wild-type sequence and removed 16 and 21 residues from the N and C termini, respectively (Methods, Extended Data Figs. 1, 2). These modifications can increase the yield of the protein, as well as its stability—both of which are crucial for the determination of the high-resolution crystal structures (Extended Data Table 1). We found that the two apo structures are essentially identical in their overall conformation at the transmembrane region, with a root-mean-square deviation (r.m.s.d.) of 1.1 Å for the C α atoms of the helix bundles (Fig. 1a, Extended Data Fig. 1). This suggests that the conformation of the receptor is not altered by crystal packing. In particular, the ECL2 region is well-folded in both structures and highly consistent (r.m.s.d. of 0.8 Å). Hereafter, we do not differentiate between these two structures and refer only to GPR52-apo unless otherwise noted. The cytoplasmic portion of GPR52-apo adopts an inactive conformation that is characterized by a lack of outward movement of transmembrane helix 6 (TM6)—due in part to the absence of binding to intracellular G proteins or other partner molecules.

ECL2 occupies the orthosteric pocket

Closer examination of the GPR52-apo structure reveals a feature that is absent in all of the other GPCR structures that are currently available. GPR52 has a 22-residue ECL2 that folds into a small module and occupies the orthosteric binding pocket of the receptor, establishing interactions with transmembrane helices to maintain its configuration (Fig. 1b, c). Structurally, ECL2 can be further divided into two segments. The first segment contains residues 179–181 and 191–200, which together function as a lid to cover the pocket from the top of the receptor (Fig. 1b, c). The second segment, in the middle of ECL2 (residues 182–190), adopts an extended conformation with a short 3_{10} helix; this segment fits into the large hydrophobic cavity in the orthosteric binding pocket and is surrounded by a panel of residues in the core of the helix (Fig. 1c). These residues are topologically equivalent in nearly all currently available class-A GPCR structures and mediate consensus contact with ligands¹⁹. Structural superposition of this second segment of ECL2 with corresponding segments from other ligand-bound GPCRs—such as the apelin receptor (APJR²⁰) in complex with the apelin peptide analogue AMG3054, and 5-HT_{2B}R²¹ in complex with the serotonin receptor agonist LSD—shows the overlay of a consensus ligand-binding pocket (Fig. 1b). In particular, on one side of this pocket, the side chain of Tyr185^{ECL2} packs tightly into a local aromatic environment formed by the residues Tyr281^{6.51}, Tyr284^{6.54} and Phe285^{6.55} of TM6 (Fig. 1c) (superscript numbers use the Ballersteros and Weinstein numbering system²²). In addition, there is a salt bridge between Lys182^{ECL2} and Asp188^{ECL2}, which appears to be a strong conformational constraint that stabilizes the second ECL2 motif in the pocket (Fig. 1c). Notably, there is a disulfide bond between Cys193^{ECL2} in the lid and Cys114^{3.25} in TM3, which fixes the conformation of the lid and in turn maintains the second ECL2 motif in the pocket (Fig. 1a). This disulfide bond is highly conserved, and has a key role in stabilizing different agonists in the orthosteric binding pocket of other class-A GPCRs¹⁹. The structural parallelism strongly suggests that the second segment of ECL2 may behave as a built-in ‘agonist’ for activating the receptor in the orthosteric binding pocket (Fig. 1b, Extended Data Fig. 3a). Consistent with this hypothesis, previous studies have shown that GPR52 exhibits high constitutive activity³. In contrast to APJR and 5-HT_{2B}R, the first half of ECL2 (residues 182–190) of GPR52 is buried in the pocket, and the second half (residues 191–199) protrudes to the extracellular surface—an ECL2 trajectory that is the

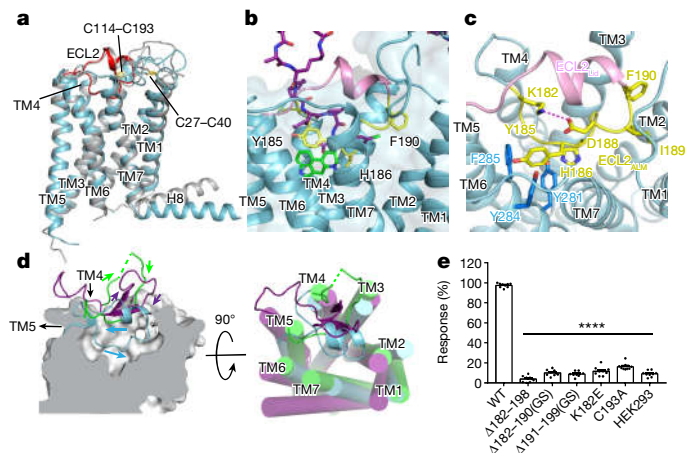


Fig. 1 | Crystal structure of GPR52-apo and analysis of ECL2. a, The GPR52-Fla-apo (grey) and GPR52-Rub-apo (blue) structures are overlaid to show the canonical seven-transmembrane-helix topology of GPR52. ECL2 is shown in red (in Rub-apo) and grey (in Fla-apo). Disulfide bonds are shown as yellow sticks. H8, helix 8. **b**, Superposition of the ECL2 ALM (yellow) with representative class-A ligands: AMG3054 in APJR (RCSB Protein Data Bank (PDB) 5VBL, purple) and LSD in 5-HT_{2B}R (PDB 5TVN, green) in the GPR52 pocket. The pocket is shown as a semitransparent surface. **c**, Magnified view of the ECL2 ALM in the orthosteric binding pocket. The lid and ALM of ECL2 are in pink and yellow cartoon representation, respectively. Key interacting residues are shown as sticks. The salt bridge between Lys182^{ECL2} and Asp188^{ECL2} is shown as a dashed line. **d**, Superposition of ECL2 in GPR52 (blue), APJR (purple) and 5-HT_{2B}R (green) in side view and top view. The ECL2 trajectories (from TM4 to TM5) are indicated with arrows. **e**, Mutations that interfere with the conformation of the ECL2 ALM in the orthosteric binding pocket abolish downstream signalling in a cAMP assay. The cAMP response level was compared between wild-type (WT) GPR52 and various mutant versions (in Δ182–190(GS) and Δ191–199(GS), residues 182–190 and 191–199, respectively, were replaced with a six-residue linker (GGSGGS)). Significance was determined by two-way analysis of variance (ANOVA) without repeated measures, followed by Dunnett’s post hoc test (**** $P < 0.0001$). Data are mean \pm s.e.m. ($n = 3$).

complete reverse of that observed in all GPCRs that have previously been described (Fig. 1d). It is also noteworthy that a conserved sodium-binding site is not present in the GPR52 structure, nor is a sodium ion observed in the orthosteric pocket—consistent with the high level of constitutive activity that is exhibited by the receptor^{18,23}.

Mutagenesis and cellular functional assays showed that deleting residues 182–198, replacing residues 182–190 or 191–199 with a 6-residue linker (GGSGGS), breaking the disulfide bond between Cys193^{ECL2} and Cys114^{3.25} or even mutating the single key residue Lys182^{ECL2} all markedly reduced the signalling activity of GPR52 (Fig. 1e). Despite these modifications, however, the levels of protein expression and membrane trafficking were maintained at 30–50% compared to the wild type (Extended Data Fig. 2d). This result confirms our hypothesis that the uniquely folded ECL2 motif has a key role in stimulating the intrinsic activity of GPR52. We refer to this new motif hereafter as the agonist-like motif (ALM) (Fig. 1c).

Cryo-EM structure of GPR52–mini-G_s

Consistent with the intrinsic activity of GPR52 and confirming our hypothesis of the ECL2 ALM, we were able to form a stable GPR52–G_s protein complex in vitro in the absence of an agonist. We purified a version of GPR52 that contained two stabilizing mutations (A130^{3.41}W and C314^{7.50}P), an N-terminal BRIL fusion protein (Methods) and an intact ICL3, and mixed it with the G_s subunit of minimal G_s (mini-G_s)^{12,24}, G_{βγ} and a camelid antibody Nb35 that binds at the G_{as}–G_β interface¹¹.

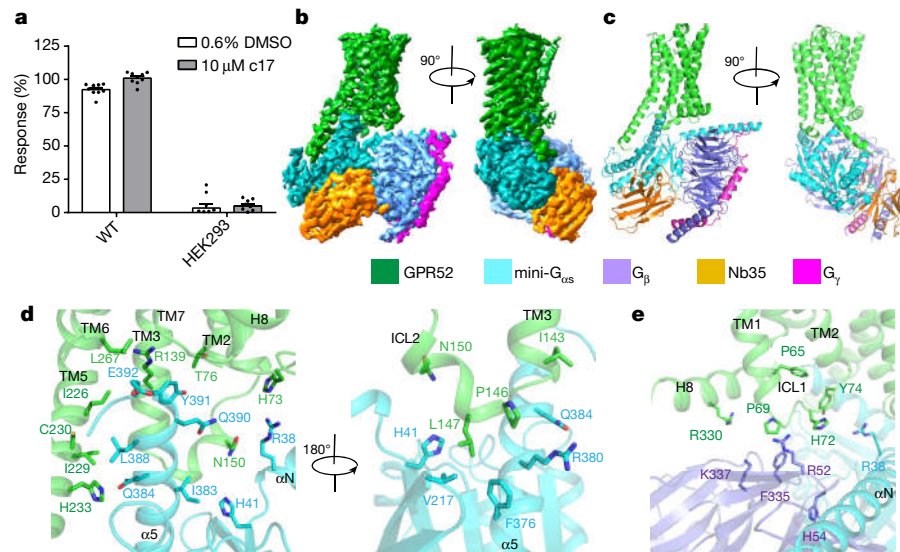


Fig. 2 | Cryo-EM structure of GPR52 bound to heterotrimeric mini-G_s in the absence of an agonist. **a**, The activity of GPR52 in the absence or presence of the c17 agonist was monitored according to cAMP response level. Data are mean \pm s.e.m. ($n = 3$). **b**, Two orthogonal views of the cryo-EM density map of the GPR52–mini-G_s–Nb35 complex, colour-coded by protein. **c**, Ribbon diagram of

the GPR52–mini-G_s–Nb35 complex in the same view as **b**. **d**, Two perspectives of the cavity formed by the cytoplasmic ends of TM2–TM3 and TM5–TM7, which accommodates the $\alpha 5$ -helix of G_{as} through both hydrophobic and electrostatic interactions. **e**, Magnified view of the interface between ICL1 of GPR52 and blade 7 of G_β. Key residues at the interface are shown as sticks.

Analysis with size-exclusion chromatography revealed that GPR52 formed a monodispersed complex with mini-G_s and Nb35 in the absence of an agonist; this complex remained stable at 4 °C for at least 10 days (Extended Data Fig. 4a). Two-dimensional classification analysis of cryo-electron microscopy (cryo-EM) data revealed averages with a clear shape that is typical of GPCR–G-protein complexes, again indicating the formation of a stable complex (Extended Data Fig. 4c). Next, we monitored the cellular activity of wild-type GPR52 in the absence or presence of c17, a surrogate agonist²⁵. In accordance with previous studies³, GPR52 showed a very high level of basal activity even in the absence of c17 (around 90% of the level of activity when c17 was present) (Fig. 2a). These results confirm our hypothesis that—unlike most of the GPCRs that have been studied so far—GPR52 can couple to heterotrimeric G_s and activate the downstream signalling pathway without stimulation from an agonist.

To understand the structural basis of the intrinsic activity of GPR52 and how this relates to G-protein coupling, we determined the cryo-EM structure of the GPR52–mini-G_s–Nb35 complex at a nominal global resolution of 3.3 Å and at around 3 Å resolution at the interface between GPR52 and mini-G_s (Fig. 2b, Extended Data Figs. 4, 5). The electron microscopy density map allowed us to unambiguously trace the polypeptide chains and build the atomic structure of the entire complex (Extended Data Table 2, Extended Data Fig. 6). The overall conformation of the GPR52–mini-G_s–Nb35 complex (hereafter referred to as GPR52–mini-G_s) is consistent with previously reported structures of class-A active GPCR–G-protein complexes^{11,12}. More specifically, the cytoplasmic region sits on the canonical surface formed by G_{as} and G_β, and Nb35 stabilizes the complex as a wedge between G_{as} and G_β (Fig. 2c).

The GPR52–mini-G_s interface consists of TM3 and TM5–TM7 of GPR52, and the Ras-like GTPase domain of G_{as} (Fig. 2d). The cytoplasmic ends of TM2–TM3 and TM5–TM7 form a deep cavity that accommodates the $\alpha 5$ helix of G_{as} through both hydrophobic and electrostatic interactions (Fig. 2d, Extended Data Fig. 7a). The aliphatic side chain of the conserved Arg139^{3,50} from TM3 of the receptor stacks with the Tyr391 aromatic ring of G_{as}, anchoring the C terminus of the $\alpha 5$ -helix of G_{as} in the centre of the cavity (Fig. 2d). TM5 has a very long cytoplasmic extension that not only forms part of the $\alpha 5$ -helix-binding cavity, but also covers a substantial area of the surface of one side of G_{as} (Fig. 2d).

On the opposite side of the $\alpha 5$ -helix-binding cavity, a short helix in ICL2 of GPR52 fits in reverse into a hydrophobic pocket that is formed by helices αN and $\alpha 5$ as well as part of the central β sheet of G_{as} (Fig. 2d, Extended Data Fig. 7b).

Compared to the structures of other GPCR–G-protein complexes^{11–16,26–32}, the interface of GPR52 with mini-G_s also involves direct interactions between G_β and ICL1 (Fig. 2e). The hydrophobic coiled ICL1 sits in a shallow groove that is formed by the seventh blade of the G_β propeller, burying an exposed surface area of around 360 Å² (Fig. 2e). The electron microscopy density map shows that Pro69^{ICL1} of GPR52 is in close proximity to, and therefore probably stacks with, the Phe335 aromatic side chain of G_β (Fig. 2e). At the periphery of the pocket, multiple charged residues of G_β are in close proximity to ICL1, suggesting that additional electrostatic contacts also contribute to the interaction between G_β and GPR52 (Fig. 2e). In addition, the side chain of Arg38 from the C-terminal end of the αN helix of G_{as} sticks out and mediates electrostatic interactions with the backbone of GPR52, which helps to anchor ICL1 in the G_β groove (Fig. 2e). Together, these results show that GPR52 mediates extensive interactions with the mini-G_s trimer and buries an exposed surface area of around 1,790 Å²—an area substantially larger than that observed in most GPCR–G complex structures^{11–16,26–32} (Fig. 2e, Extended Data Fig. 7c–h).

From apo to mini-G_s-coupled states

A comparison of the apo and the mini-G_s-coupled structures of GPR52 enables us to examine the conformational changes that are associated with complex formation. In the GPR52–apo structure, the cytoplasmic portion of the long TM5 was modified to facilitate the crystallization of the protein (Fig. 3a). Structural overlay reveals an outward movement of TM6 of about 6 Å (based on C α of Val266^{6,36}) and a seesaw winding of TM5 (Fig. 3b) in the mini-G_s-coupled relative to the apo structure. The latter is characterized by an N-terminal inward movement towards TM4 and a C-terminal outward movement towards TM6 in mini-G_s-coupled GPR52 (Fig. 3b). These movements induce a rotamer change of the key amino acid Arg139^{3,50}, pointing it towards the transmembrane core; as a result, TM3 shifts upwards by around 1.5 Å (based on C α of Arg139^{3,50}) and Arg139^{3,50} stacks on the C terminus of the G_{as} $\alpha 5$ helix in the complex

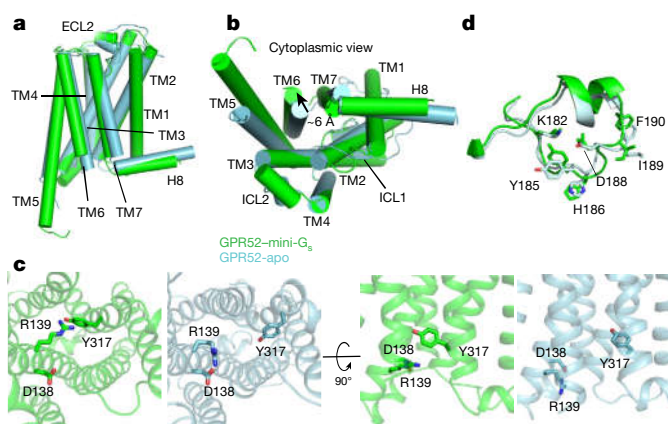


Fig. 3 | Structural comparison of apo and mini- G_s -coupled GPR52. **a**, Side view of the overlaid mini- G_s -coupled (green) and apo (blue) GPR52 structures. **b**, Cytoplasmic view of the structural comparison reveals an outward movement of TM6 of around 6 Å in the GPR52–mini- G_s relative to the GPR52–apo structure. **c**, Two magnified orthogonal views of the conformational changes in R139^{3,50} and Y317^{7,53}, as well as the peripheral transmembrane helices, between the mini- G_s -coupled (green) and apo (blue) GPR52 structures. D138^{3,49}, R139^{3,50} and Y317^{7,53} are shown as sticks. **d**, Superposition of ECL2 from the GPR52–apo crystal structure (blue) and GPR52–mini- G_s cryo-EM structure (green), showing the conformational consistency between the structures.

(Fig. 3c). The ‘ionic lock’ between residues Arg^{3,50} and Glu^{6,30} or Asp^{6,30} is a conserved conformational signature of all inactive GPCRs^{33–36}. Although Asp260^{6,30} is not visible in the crystal structure of GPR52–apo, the salt bridge between Arg139^{3,50} and Asp138^{3,49} indicates that GPR52–apo is not in an active conformation (Fig. 3c). In addition, during the transition from the apo to the mini- G_s -coupled state, the side chain of another conserved residue, Tyr317^{7,53}, rotates into the core of the helix bundle to lock TM6 into the active position for G-protein binding (Fig. 3c). Of note, the C terminus of TM6 is partially disordered in the electron microscopy map, suggesting that this region is structurally dynamic upon G-protein binding. Together, these structural rearrangements open space in the helical core at the cytoplasmic region of GPR52, generating a crevice of a sufficient size to accommodate the C terminus of the α_5 helix of G_{as} .

In contrast to the cytoplasmic region, the extracellular end of GPR52 that forms the canonical orthosteric binding pocket does not exhibit marked changes between the apo and the mini- G_s -coupled GPR52 structures (Fig. 4f). In particular, the conformation of the ECL2 region is highly conserved, with a C α atom r.m.s.d. of 1.2 Å (Fig. 3d, Extended Data Fig. 6b). This finding suggests that the association with mini- G_s induces major structural changes only in the cytoplasmic region of the receptor. Notably, this closely resembles the changes in conformation that occur during the transition of receptors A_{2A}R and 5-HT_{1B}R from the agonist-bound inactive state to the active G-protein-coupled state^{12,26} (Extended Data Fig. 3c–e). In both cases, coupling with G proteins induces conformational changes only on the cytoplasmic side of the agonist-bound receptors that are capable of transducing downstream signalling^{12,26} (Extended Data Fig. 3c, d). Comparison of GPR52 to the A_{2A}R and 5-HT_{1B}R structures in the agonist-bound inactive state shows that structural features at the activation switches—such as the DRY motif—are highly consistent (Extended Data Fig. 3e). Given that GPR52 can bind the trimeric G_s protein and exhibits a high level of constitutive activity in cells without an agonist³ (Fig. 2a), it is likely that the ECL2 ALM is functionally equivalent to an agonist in other GPCRs.

Crystal structure of ligand-bound GPR52

The canonical orthosteric pocket in GPR52 is almost fully occupied by ECL2, leaving almost no space for a new ligand to bind. We therefore set

out to investigate the binding mode of a GPR52 agonist by cocrystallizing GPR52 with the surrogate ligand c17. We used the same construct here as for the crystallographic study of GPR52–Fla–apo. The resulting structure was determined at 2.2 Å resolution. The overall conformation of the c17-bound GPR52 is highly consistent with that of GPR52–apo, with a C α atom r.m.s.d. of 1.7 Å (Extended Data Fig. 1b). The agonist c17 is in a C-shaped configuration perpendicular to the membrane plane, and sits in the pocket through shape complementarity (Fig. 4a, b). The pocket, which is mainly formed by TM1, TM2, TM7 and ECL2, is very close to the extracellular surface (Fig. 4b, c). This is in contrast to all available structures of GPCRs in complex with ligands, in which TM1 is not in direct contact with any of the ligands¹⁹. In the GPR52–c17 complex, the N-terminal loop and ECL2 push the ligand towards one side and contribute to the formation of a new ligand pocket, which we refer to as the ‘side pocket’ to distinguish it from the canonical orthosteric binding site. The space in the side pocket is substantially limited by ECL2 and TM7; at the narrowest point it is only 3.7 Å wide, which would barely fit a single layer of planar aromatic rings.

The interactions between GPR52 and c17 involve hydrogen bonds, hydrophobic contacts and aromatic stacking (Fig. 4b). There are four pairs of hydrogen bonds between the ligand and the main-chain atoms of the receptor: the top hydroxyl group of c17 with Ile189^{ECL2} and Glu191^{ECL2}; the middle hydroxyl group with Asp188^{ECL2}; and the amide group with Cys40^{1,32}. The ligand is further stabilized in the side pocket through contacts with a group of hydrophobic residues (Tyr34^{N-term}, Val39^{1,31}, Ile47^{1,39}, Leu101^{2,64}, Phe117^{3,28} and Thr303^{7,39}). In particular, the aromatic ring system of c17 forms π – π interactions with Phe300^{7,36} in the bottom of the pocket (Fig. 4b). Notably, molecular docking analysis suggests that all reported GPR52 agonists^{6,7,25} probably occupy the side pocket in a similar manner, although with different chemical scaffolds (Extended Data Fig. 8a).

The side pocket in the structures of the GPR52–apo and GPR52–mini- G_s complexes is smaller but consistent overall with that in the GPR52–c17 structure, except for the N-terminal loop, which is highly flexible in the apo structures without the constraint of a bound ligand (Fig. 4c). We noted that the Trp304^{7,40} indole ring rotates into the bottom of the pocket core in the apo structure, reducing the size of the lower part of the pocket that would otherwise clash with the c17 ligand (Fig. 4c). Substituting Trp304^{7,40} with an alanine residue did not influence the c17-induced signalling activity (Fig. 4d), consistent with the Trp304^{7,40} side chain not interacting with c17 in the ligand-bound state. At the same time, the pocket residues His25 and Ser26 on the N terminus, Ile189 on ECL2 and Phe117^{3,28} on TM3 extend closer to the pocket core in the apo compared to the GPR52–c17 structure (Fig. 4c), further shrinking the side pocket. Such a small pocket would bind only to ligands of a specific shape and size. The overall conformation of ECL2 in the GPR52–c17 structure, however, is consistent with that of the GPR52–apo and GPR52–mini- G_s structures (Fig. 4e, f), suggesting that the ECL2 conformation is not altered by the side-pocket occupancy.

Unique pocket and receptor conformation

Knowledge of this small side pocket provides valuable insight in the search for tool compounds and an endogenous ligand for GPR52. We compared the side pocket of GPR52 to that of representative peptide receptors, non-lipid small-molecule receptors and lipid-activated receptors. We found that the side pocket is markedly smaller in volume in GPR52 than in peptide receptors (1,054 Å³ for c17-bound GPR52 and 958 Å³ for GPR52–apo) (see Supplementary Table 2 for the comparison with other receptors).

As GPR52 is colocalized with D2R in the basal ganglia and activation of GPR52 counteracts signalling from the $G_{i/o}$ -coupled D2R¹, we compared the structure of GPR52–c17 with that of D2R as a representative small-molecule receptor³⁷ (Fig. 5a). We found that whereas c17 in GPR52 is located closer to TM1, TM2 and TM7, the D2R ligand (risperidone)

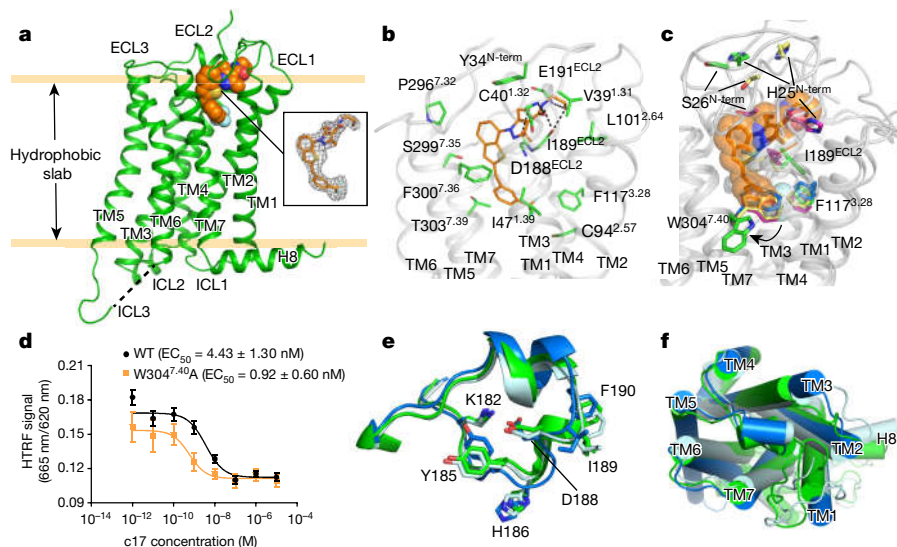


Fig. 4 | Crystal structure of GPR52–c17 and a novel ligand-binding pocket.

a, The structure of the GPR52–c17 complex (green) with ligand c17 (orange) in the binding pocket. ICL3 is shown as a dashed line. The $F_o - F_c$ omit map (insert, grey mesh, contoured at 3.0 σ) shows clear electron density for ligand c17. **b**, Key molecular interactions in the c17-binding pocket. c17 (orange) and the GPR52 residues that are involved in ligand binding (green) are shown as sticks, the receptor is in grey and hydrogen bonds are black dashed lines. **c**, Comparison of the ligand-binding pocket between GPR52-*apo* and GPR52–c17. Residues in the *apo* structure (purple for Rub-*apo*; yellow for Fla-*apo*), in GPR52–c17 (green) and in GPR52–mini- G_s (blue) are shown. Ligand c17 is shown

as orange sticks overlaid with a transparent sphere representation. The four structures are superimposed and shown in grey cartoon. **d**, Assessment of the efficacy of c17 as an agonist in wild-type GPR52 and GPR52 with a mutated residue in the pocket (W304⁷⁴⁰A). Data are mean \pm s.e.m. ($n = 3$). HTRF, homogeneous time-resolved fluorescence. EC₅₀, half maximal effective concentration. **e**, ECL2 of GPR52-*apo* (cyan), GPR52–c17 (green) and GPR52–mini- G_s (blue) are superimposed. Key residues are shown as sticks. **f**, Extracellular view of superimposed GPR52-*apo* (cyan), GPR52–c17 (green) and GPR52–mini- G_s (blue), showing the consistent ECL2 conformation in the orthosteric binding pocket.

is closer to TM4–TM6—a ligand-binding mode that is more commonly seen in other small-molecule³⁸ and lipid receptors³⁹ (Fig. 5b, Supplementary Table 3). The arrangement of the conserved proline residue in TM5 of GPR52 is unusual in that Pro214^{5.50} overlaps with Val^{5.49} but not the equivalent Pro^{5.50} in D2R; this confers a winding mode to TM5 that is unique to GPR52 among regular class-A GPCR structures (Fig. 5a, c).

A structural similarity network for all class-A GPCRs for which the inactive structures have been reported showed that GPR52 was the only receptor with a C α atom r.m.s.d. of more than 2.0 Å based on the conformations of the transmembrane helices (Extended Data Fig. 8b). The unique position of Pro214^{5.50} and the resulting winding mode of TM5 are seen in all GPR52 structures, including the active-state GPR52–mini- G_s complex. This configuration indicates that GPR52 cannot use the consensus P^{5.50}-I^{3.40}-F^{6.44} (PIF) motif³¹ to trigger the outward bending of TM6 during activation (Extended Data Fig. 3b), and thus may have to use a different mechanism of G-protein coupling.

Finally, the closest homologue of GPR52 is GPR21, another orphan receptor. With overall sequence identity as high as 71% (Extended Data Fig. 8c), the two orphan receptors are especially consistent at ECL2. However, the ligand-binding residues in the side pocket are less conserved, with only 59% identity. The less-conserved side pocket suggests the possibility of distinct recognition modes and thus the potential for development of selective modulators.

Discussion

The structure of GPR52 that we present here will allow accurate modelling of other homologous orphan receptors. Our observation that ECL2 occupies the orthosteric binding pocket and pushes ligand c17 into a very narrow cavity between ECL2 and TM1, TM2 and TM7 on one side suggests that c17 may function as an allosteric agonist and fit into a side pocket that has not been observed in the structures of other GPCR–ligand complexes or allosteric modulators. Analysis of the shape and the hydrophobic nature of the side pocket may provide

further insight that will aid in the deorphanization of GPR52. Given the small volume and hydrophobicity of the side pocket in GPR52, we propose that the endogenous ligand (if any) that binds to this pocket is most likely to be a small lipid molecule, as previously predicted by

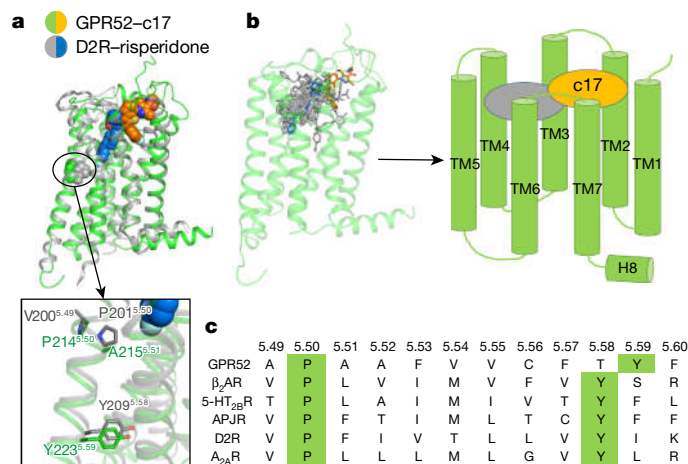


Fig. 5 | Unique ligand-binding mode and structural conformation of GPR52 compared to other class-A GPCRs.

a, Side view of the GPR52–c17 complex (green–orange) and D2R–risperidone complex (grey–blue). The conserved Pro^{5.50} residues in GPR52 and D2R are shown as sticks. **b**, Position comparison of c17 in GPR52 with class-A small-molecule receptors. Representative agonist ligands from 27 structures of GPCR–ligand complexes (grey sticks) are superimposed on the GPR52 receptor structure (green) with c17 (orange) and risperidone (blue). The schematic on the right shows the binding sites of common ligands (grey) and c17 (orange) in side view. **c**, Sequence alignment of part of TM5 in GPR52, β_2 AR, 5-HT_{2B}R, APJR, D2R and A_{2A}R. Conserved residues are coloured in green. The top labels indicate Ballersteros and Weinstein numbering²².

phylogeny analysis⁴⁰. Therefore, as GPR52 is primarily expressed in the brain, we suggest that future deorphanization efforts should start by screening small lipids that are found in brain tissues.

We have demonstrated that ECL2 can serve as a built-in agonist, leading to an active conformation of GPR52 and a high level of basal signalling. This further highlights the need to search for an antagonist or inverse agonist for research and therapeutic applications, but thus far only one low-efficacy antagonist has been reported¹⁰. The GPR52 ligand-binding side pocket we have revealed can be targeted by rational structure-based ligand design and holds promise for selective drug screening owing to its allosteric-like features.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2019-0>.

- Yao, Y. et al. A striatal-enriched intronic GPCR modulates huntingtin levels and toxicity. *eLife* **4**, e05449 (2015).
- Komatsu, H. et al. Anatomical transcriptome of G protein-coupled receptors leads to the identification of a novel therapeutic candidate GPR52 for psychiatric disorders. *PLoS ONE* **9**, e90134 (2014).
- Martin, A. L., Steurer, M. A. & Aronstam, R. S. Constitutive activity among orphan class-A G protein coupled receptors. *PLoS ONE* **10**, e0138463 (2015).
- Foster, S. R. et al. Discovery of human signaling systems: pairing peptides to G protein-coupled receptors. *Cell* **179**, 895–908 (2019).
- Krutzke, S. K. et al. Array-based molecular karyotyping in fetal brain malformations: identification of novel candidate genes and chromosomal regions. *Birth Defects Res. A Clin. Mol. Teratol.* **106**, 16–26 (2016).
- Setoh, M. et al. Discovery of the first potent and orally available agonist of the orphan G-protein-coupled receptor 52. *J. Med. Chem.* **57**, 5226–5237 (2014).
- Nishiyama, K. et al. FTBMT, a novel and selective GPR52 agonist, demonstrates antipsychotic-like and procognitive effects in rodents, revealing a potential therapeutic agent for schizophrenia. *J. Pharmacol. Exp. Ther.* **363**, 253–264 (2017).
- Komatsu, H. Novel therapeutic GPCRs for psychiatric disorders. *Int. J. Mol. Sci.* **16**, 14109–14121 (2015).
- Alavi, M. S., Shamsizadeh, A., Azhdari-Zarmehri, H. & Roohbakhsh, A. Orphan G protein-coupled receptors: the role in CNS disorders. *Biomed Pharmacother.* **98**, 222–232 (2018).
- Song, H. et al. Targeting Gpr52 lowers mutant HTT levels and rescues Huntington's disease-associated phenotypes. *Brain* **141**, 1782–1798 (2018).
- Rasmussen, S. G. et al. Crystal structure of the β_2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
- García-Nafria, J., Lee, Y., Bai, X., Carpenter, B. & Tate, C. G. Cryo-EM structure of the adenosine A_{2A} receptor coupled to an engineered heterotrimeric G protein. *eLife* **7**, e35946 (2018).
- Liang, Y. L. et al. Phase-plate cryo-EM structure of a class B GPCR–G-protein complex. *Nature* **546**, 118–123 (2017).
- Zhang, Y. et al. Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* **546**, 248–253 (2017).
- Liang, Y. L. et al. Cryo-EM structure of the active, G_s-protein complexed, human CGRP receptor. *Nature* **561**, 492–497 (2018).
- Zhao, L. H. et al. Structure and dynamics of the active human parathyroid hormone receptor-1. *Science* **364**, 148–153 (2019).
- Rausser, L., Savage, J. E., Meltzer, H. Y. & Roth, B. L. Inverse agonist actions of typical and atypical antipsychotic drugs at the human 5-hydroxytryptamine_{2C} receptor. *J. Pharmacol. Exp. Ther.* **299**, 83–89 (2001).
- Katritch, V. et al. Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39**, 233–244 (2014).
- Venkatakrishnan, A. J. et al. Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
- Ma, Y. et al. Structural basis for apelin control of the human apelin receptor. *Structure* **25**, 858–866 (2017).
- Wacker, D. et al. Crystal structure of an LSD-bound human serotonin receptor. *Cell* **168**, 377–389 (2017).
- Ballesteros, J. A. & Weinstein, H. in *Receptor Molecular Biology, Methods in Neurosciences* vol. 25 (ed. Sealfon, S. C.) 366–428 (Elsevier, 1995).
- Wang, S. & Wacker, D. D₄ dopamine receptor high-resolution structures enable the discovery of selective agonists. *Science* **358**, 381–386 (2017).
- Carpenter, B. & Tate, C. G. Engineering a minimal G protein to facilitate crystallisation of G protein-coupled receptors in their active conformation. *Protein Eng. Des. Sel.* **29**, 583–594 (2016).
- Nakahata, T. et al. Design and synthesis of 1-(1-benzothiophen-7-yl)-1H-pyrazole, a novel series of G protein-coupled receptor 52 (GPR52) agonists. *Bioorg. Med. Chem.* **26**, 1598–1608 (2018).
- García-Nafria, J., Nehmé, R., Edwards, P. C. & Tate, C. G. Cryo-EM structure of the serotonin 5-HT_{1B} receptor coupled to heterotrimeric G_s. *Nature* **558**, 620–623 (2018).
- Kang, Y. et al. Cryo-EM structure of human rhodopsin bound to an inhibitory G protein. *Nature* **558**, 553–558 (2018).
- Koehl, A. et al. Structure of the μ -opioid receptor–G_i protein complex. *Nature* **558**, 547–552 (2018).
- Maeda, S., Qu, Q., Robertson, M. J., Skiniotis, G. & Kobilka, B. K. Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science* **364**, 552–557 (2019).
- Qi, X. et al. Cryo-EM structure of oxysterol-bound human Smoothed coupled to a heterotrimeric G_s. *Nature* **571**, 279–283 (2019).
- Draper-Joyce, C. J. et al. Structure of the adenosine-bound human adenosine A_1 receptor–G_i complex. *Nature* **558**, 559–563 (2018).
- Kato, H. E. et al. Conformational transitions of a neurotensin receptor 1–G_{i1} complex. *Nature* **572**, 80–85 (2019).
- Ballesteros, J. A. et al. Activation of the β_2 -adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* **276**, 29171–29177 (2001).
- Greasley, P. J., Fanelli, F., Rossier, O., Abuin, L. & Cotecchia, S. Mutagenesis and modelling of the α_{1B} -adrenergic receptor highlight the role of the helix 3/helix 6 interface in receptor activation. *Mol. Pharmacol.* **61**, 1025–1032 (2002).
- Shapiro, D. A., Kristiansen, K., Weiner, D. M., Kroeze, W. K. & Roth, B. L. Evidence for a model of agonist-induced activation of 5-hydroxytryptamine 2A serotonin receptors that involves the disruption of a strong ionic interaction between helices 3 and 6. *J. Biol. Chem.* **277**, 11441–11449 (2002).
- Romo, T. D., Grossfield, A. & Pitman, M. C. Concerted interconversion between ionic lock substates of the β_2 adrenergic receptor revealed by microsecond timescale molecular dynamics. *Biophys. J.* **98**, 76–84 (2010).
- Wang, S. et al. Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature* **555**, 269–273 (2018).
- Munk, C. et al. An online resource for GPCR structure determination and analysis. *Nat. Methods* **16**, 151–162 (2019).
- Audet, M. & Stevens, R. C. Emerging structural biology of lipid G protein-coupled receptors. *Protein Sci.* **28**, 292–304 (2019).
- Kakarala, K. K. & Jamil, K. Sequence-structure based phylogeny of GPCR class A rhodopsin receptors. *Mol. Phylogenet. Evol.* **74**, 66–96 (2014).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Molecular cloning of GPR52 crystallization constructs

The human *GPR52* gene was subcloned into the expression vector pFast-bac1. For GPR52-Fla-apo, the flavodoxin⁴¹ fusion protein was inserted into ICL3, replacing residues 236–261 of GPR52. The construct contained residues 17–340 of GPR52 and a 10×His-tag at the C terminus that is removable by cleavage with 3C protease. Haemagglutinin (HA) signal peptide, Flag tag and thermostabilized *Escherichia coli* apocytochrome *b₅₆₂*RIL (BRIL)⁴¹ were added on the N terminus to enhance receptor expression and were then removed during purification with a tobacco etch virus (TEV) site. To improve protein stability, seven point mutations were introduced: A130^{3,41}W (ref. ⁴²), A264^{6,34}L, W278^{6,48}Q, C314^{7,50}P, S318^{7,54}A, N321^{8,47}D and V323^{8,49}T. Cloning of the ligand-free GPR52-Rub-apo construct was performed as described for GPR52-Fla-apo, except that the rubredoxin⁴¹ fusion protein was inserted into ICL3 (replacing residues 235–263 of GPR52).

Expression and purification of GPR52 crystallization constructs

The GPR52 protein was expressed and purified as previously described⁴³. We used the Bac-to-Bac baculovirus system (Invitrogen) in *Spodoptera frugiperda* (Sf9) cells for expression. These cells were infected with baculovirus at a density of 2×10^6 cells per ml. Cells were grown at 27 °C and collected 48 h after infection.

For cocrystallization with ligand c17, the cell membrane was washed with a low-salt buffer (10 mM HEPES (pH 7.5), 20 mM KCl, 10 mM MgCl₂ and protease inhibitor cocktail (Roche)) and three times with a high-salt buffer (10 mM HEPES (pH 7.5), 1 M NaCl, 20 mM KCl, 10 mM MgCl₂ and protease inhibitor cocktail). Before solubilization, purified membranes were incubated with 20 μM c17 for 3 h, then incubated with 2 mg ml⁻¹ iodoacetamide (Sigma) at 4 °C for 1 h. The protein was extracted from the membrane by 100 mM HEPES, 800 mM NaCl, 1.0% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM) (Anatrace) and 0.2% (w/v) cholesteryl hemisuccinate (CHS) (Sigma) and stirred for 2.5 h at 4 °C. After centrifugation, the supernatant was incubated with TALON IMAC resin (Clontech) at 4 °C overnight. Then the resin was washed with 15 column volumes of buffer I: 50 mM HEPES (pH 7.5), 800 mM NaCl, 5% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS, 10 mM MgCl₂, 20 mM imidazole and 20 μM c17. The resin was resuspended with 3 column volumes of buffer II: 25 mM HEPES (pH 7.5), 800 mM NaCl, 5% (v/v) glycerol, 0.03% (w/v) DDM, 0.006% (w/v) CHS, 40 mM imidazole and 20 μM c17. TEV protease was added with a molar ratio of 1:10, and the mixture was incubated at 4 °C overnight. Next, the resin was washed with 8 column volumes of buffer II to remove the HA, Flag tag and BRIL. The protein was eluted with 3 column volumes of buffer III: 25 mM HEPES (pH 7.5), 800 mM NaCl, 5% (v/v) glycerol, 0.01% (w/v) DDM, 0.002% (w/v) CHS, 220 mM imidazole and 20 μM c17. The ligand-free GPR52 protein was purified as described above, except that ligand c17 was not added during the purification process.

Crystallization

The receptor was concentrated to around 50 mg ml⁻¹ with a 100-kDa cut-off concentrator (Millipore). The protein sample was reconstituted into a lipidic cubic phase by mixing 40% protein with 60% lipid (10% (w/w) cholesterol and 90% (w/w) monoolein) in a syringe mixer. The crystallization trials were set up by crystallization robot NT8 (Formulatrix). The protein and lipid mixture was dispensed in 40-nl volumes on 96-well glass sandwich plates and overlaid with 800 nl of precipitant solution in each well. For the GPR52–c17 complex, crystals appeared after 1 day in 0.13–0.18 M sodium acetate, 0.1 M sodium citrate pH 5.0

and 32–35% PEG400, and reached full size ($40 \times 40 \times 50 \mu\text{m}^3$) after 9 days.

For GPR52-Fla-apo, crystals appeared after 1 day in 0.1 M potassium acetate, 0.1 M sodium citrate pH 5.0 and 30% PEG400, and reached full size ($15 \times 15 \times 80 \mu\text{m}^3$) after 5 days. For GPR52-Rub-apo, crystals appeared after 1 day in 0.08–0.1 M magnesium sulfate, 0.1 M sodium cacodylate trihydrate pH 6.2 and 28–31% PEG300, and reached full size ($120 \times 60 \times 10 \mu\text{m}^3$) after 1 week.

Crystal data collection and structure determination

The X-ray diffraction data were collected at SPring-8 beamline 41XU, Hyogo, Japan, using an EIGER X 16M detector (X-ray wavelength 1.0000 Å). A rastering system was used to find the best diffracting region of single crystals⁴⁴. The crystals were exposed with a $10 (\text{v}) \times 9 (\text{h})$ -μm minibeam for 0.2 s and 0.2° oscillation per frame. All three datasets (GPR52-Fla-apo, GPR52-Rub-apo and GPR52–c17) were processed using HKL2000⁴⁵. For GPR52-Fla-apo, a molecular replacement method with Phaser⁴⁶ was used to obtain the initial phase information, using the structures of a homologous model of human GPR52 generated by Rosetta⁴⁷ and flavodoxin (PDB 1II0) as search models. The initial model and map were subjected to PHENIX AutoBuild using all data including weak reflections of lower than 2.9 Å resolution. Most of the transmembrane helices of GPR52 were traced. Refinement was performed with PHENIX⁴⁸ and REFMAC5⁴⁹, followed by manual examination and rebuilding of the refined coordinates in Coot⁵⁰ using both $2F_o - F_c$ and $F_o - F_c$ maps. Both the GPR52-Fla-apo and GPR52-Rub-apo datasets showed strong anisotropy and were truncated on the STARANISO server (<http://staraniso.globalphasing.org>) using the diffraction CC_{1/2} cut-off criterion of 0.10. The GPR52-Rub-apo structure was solved by Phaser with the truncated model of GPR52-Fla-apo by deleting the flavodoxin part. The rubredoxin molecule taken from PDB entry 6BD4 served as a second searching model by Phaser to generate a complete model of GPR52-Rub-apo. The GPR52-Fla–c17 structure was solved by molecular replacement with GPR52-Fla-apo as the searching model after removing all water and other molecules. The model and restraints of c17 were created using the Ligand Builder in Coot and placed into the electron density map in the orthosteric binding pocket. The Ramachandran plot determined by MolProbity⁵¹ indicates that 96.1% (3.9%) of residues in GPR52-Fla-apo, 98.3% (1.7%) of residues in GPR52-Rub-apo and 98.2% (1.8%) of residues in GPR52–c17 were in favoured (allowed) regions. Data collection and structure refinement statistics are listed in Extended Data Table 1.

Purification and formation of the GPR52–mini-G_s–Nb35 complex

For cryo-EM research, the construct of human GPR52 (residues 1–340) was designed with two point mutations (A130^{3,41}W and C314^{7,50}P) and an N-terminal BRIL fusion protein. Up to the point of the buffer II wash, the purification process was the same as for ligand-free GPR52. After the resin was washed with 8 column volumes of buffer II, the wash buffer was changed to 3 column volumes of exchange buffer: 50 mM HEPES (pH 7.5), 500 mM NaCl, 10% (v/v) glycerol, 0.5% (w/v) lauryl maltose neopentyl glycol (LMNG) (Anatrace), 0.1% (w/v) CHS, 10 mM MgCl₂ and 20 mM imidazole and incubated for 4 h at 4 °C before HRV 3C protease (with a ratio 1:20) was added to cleave the 10×His-tag. The flow-through was collected and 3 column volumes of exchange buffer were added and collected. The protein solution was concentrated to a volume of 0.5 ml and loaded onto a Superdex 200 10/300 column (GE Healthcare) equilibrated with 20 mM HEPES (pH 7.5), 100 mM NaCl, 0.00075% (w/v) LMNG and 0.00015% (w/v) CHS. Peak fractions corresponding to GPR52 were pooled and concentrated to 20 mg ml⁻¹.

The G_{αs} subunit of mini-G_s (mini-G_{αs}) used in this study was the same as that used in a previous study of the cryo-EM structure of A_{2A}R–mini-G_s–Nb35¹². In brief, mini-G_{αs} was expressed in the *E. coli* strain BL21 and purified by Ni²⁺ affinity chromatography, followed by cleavage of the

His-tag using TEV protease and negative purification on Ni²⁺-NTA agarose to remove the TEV and undigested mini-G_{as}. The G_{by} and Nb35 purification was performed following previously described protocols^{12,24}.

Heterodimeric G_{by} and mini-G_{as} were mixed in a 1:1.2 ratio and incubated on ice for 4 h to form heterotrimeric mini-G_s. Excess mini-G_{as} was removed by size-exclusion chromatography on a Superdex 200 10/300 column with a running buffer of 20 mM HEPES (pH 7.5), 100 mM NaCl, 10% glycerol, 1 mM MgCl₂, 1 μM GDP and 0.1 mM Tris(2-carboxyethyl) phosphine (TCEP).

Purified GPR52 was mixed with a 1.2-fold molar excess of heterotrimeric mini-G_s and 1.5-fold molar excess of Nb35 in the presence of apyrase (0.2 U ml⁻¹). The mixture was incubated on ice overnight. The sample was loaded on to a Superdex 200 10/300 column. Peak fractions containing the GPR52–mini-G_s–Nb35 complex were pooled and concentrated to 4 mg ml⁻¹.

Preparation of vitrified sample

A 3-μl droplet of purified sample at a concentration of around 2.5 mg ml⁻¹ was applied to glow-discharged holey carbon grids (Quantifoil, 200 mesh copper R1.2/1.3). Excess sample was removed by blotting with filter paper for 3.5 s before plunge-freezing in liquid ethane using a FEI Vitrobot Mark IV at 100% humidity and 4 °C.

Cryo-EM data acquisition

Images were collected on a FEI Titan Krios microscope at 300 kV using a Falcon III detector in electron-counting mode. EPU software (FEI) was used for automatic data collection. Data were collected in three independent sessions to give a total of 7,287 movies. Each stack was exposed for 53 s with a dose of 0.95 e⁻ per pixel per second, resulting in a total of 32 frames per stack. The total dose rate was around 40 e⁻ Å⁻² for each stack.

Cryo-EM data processing and model building

Drift correction and dose weighting were performed on the 7,287 movies with MotionCor2⁵². The contrast transfer function (CTF) parameters were estimated by Gctf⁵³ and all three-dimensional (3D) reconstructions were performed with RELION-3⁵⁴. Around 2,000 particles were manually picked to generate the templates for automatic picking of particles. After automatic picking, 8,861,544 particles were extracted using a box size of 256 pixels. Next, false positives or 'bad particles' were eliminated over two rounds of reference-free two-dimensional (2D) classification. A total of 4,849,328 particles were finally selected and subjected to a global angular search 3D classification using a 60 Å low-pass-filtered initial model that was generated from the A_{2A}R–G_s complex¹² (PDB 6GDG) with 1 class and 25 iterations. The outputs of the 1st to 25th iterations were subjected to local angular search 3D classification with 5 separate classes. After merging of all good classes and removal of duplicated particles, 1,551,107 particles were subjected to 3D autorefinement. Then the particles were used for further 3D classification with no alignment, and 651,456 particles were selected for further autorefinement. After CTF refinement and Bayesian polishing, final 3D refinement resulted in an overall structure at 3.32-Å resolution. All resolutions were based on the gold standard (two halves of data refined independently) Fourier shell correlation (FSC) = 0.143 criterion⁵⁵. RELION-3 was used to estimate the variations in local resolution of the density map.

The crystal structures of human GPR52–apo (PDB 6LI2) and the G-protein complex taken from A_{2A}R bound to the mini-G_s heterotrimer (PDB 6GDG) were used as initial models for model rebuilding and refinement against the electron microscopy map. All models were docked into the electron microscopy density map using UCSF Chimera⁵⁶, followed by iterative manual adjustment in Coot⁵⁷, fragment-based refinement with Rosetta⁴⁷ and real-space refinement using PHENIX⁴⁸. The model statistics were validated using MolProbity⁵¹. Structural figures were prepared in UCSF Chimera and PyMOL (<https://pymol.org/2/>). Model overfitting was evaluated by refinement against one cryo-EM half map.

FSC curves were calculated on the basis of the final GPR52–mini-G_s–Nb35 model and the half map that was used for refinement, as well as the other half map that was used for cross-validation.

GPR52–G_s-mediated cAMP assay

The wild-type *GPR52* gene was subcloned in vector pcDNA3.0 with an N-terminal HA signal peptide and Flag tag. Mutations were introduced by QuickChange PCR. HEK293T cells were cultured in 1× DMEM supplemented with 10% (v/v) fetal bovine serum and incubated in 5% CO₂ at 37 °C. Before the transfection, cells were seeded on 6-cm cell-culture plates. When cells had grown to approximately 2 × 10⁶, wild-type or mutant *GPR52* DNA was transfected into cells using Lipofectamine 2000 reagent (Life Technologies). After a 24-h culture, cells were collected and resuspended in PBS containing 500 μM IBMX at a density of 2 × 10⁵ cells per ml. Cells were then plated onto 384-well assay plates at 1,000 cells per 5 μl per well. Another 5 μl of buffer containing c17 at various concentrations was added to the cells, and they were incubated for 30 min incubation at 37 °C. Intracellular cAMP measurement was carried with a Cisbio HTRF Dynamic 2 cAMP kit (Cisbio) and an EnVision multi-plate reader according to the manufacturer's instructions⁵⁸. The HTRF ratio was converted to a response (%) using the following formula: response(%) = (ratio_{sample} – ratio_{max}) / (ratio_{min} – ratio_{max}) × 100. Cell-surface expression for each mutant was monitored by a fluorescence-activated cell sorting (FACS) assay. In brief, the expressed cells were incubated with mouse anti-Flag (M2–fluorescein isothiocyanate (FITC)) antibody (Sigma) for 20 min at 4 °C, and then a 9-fold excess of PBS was added to cells. Finally, the surface expression of GPR52 was monitored by detecting the fluorescent intensity of FITC using a Guava EasyCyte HT system (Millipore).

Construction of structural similarity network

Class-A GPCRs for which the inactive structure has been reported (Supplementary Table 4) were used to construct the similarity network. For receptors with more than one reported inactive structure, only one structure was selected (receptor state, completeness and resolution were considered). The r.m.s.d. values between C_α atoms in the transmembrane helices were calculated using UCSF Chimera for edge attributes. In total, 192 residues were included: Ballersteros and Weinstein numbers 1.35–57, 2.37–63, 3.22–56, 4.39–63, 5.36–65, 6.33–59 and 7.31–55. The visualization of the network was generated with Cytoscape software⁵⁹.

Structure and sequence comparison

The calculation of the pocket volume was performed with the program Cavity⁶⁰. The representation of the GPR52 and GPR21 sequence alignment was generated using the ESPript website⁶¹ (<http://esript.ibcp.fr>).

Molecular docking

Molecular docking was performed using Schrödinger software. Processing of the protein structure was performed with the Protein Preparation Wizard (<https://www.schrodinger.com/protein-preparation-wizard>); conversion of ligands from 2D to 3D structures was performed using LigPrep (<https://www.schrodinger.com/ligprep>); and docking was performed with Glide 6.9 (<https://www.schrodinger.com/gleide>) in standard precision. The cartoons of all structures were generated by PyMOL⁶².

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The coordinates and structure factors for GPR52–Fla–apo, GPR52–Rub–apo, GPR52–Fla–c17 and GPR52–mini-G_s–Nb35 have been deposited in

the PDB with accession codes 6LI1, 6LI2, 6LI0 and 6LI3, respectively. The cryo-EM 3D maps of the GPR52–mini-G_s–Nb35 complex have been deposited in the Electron Microscopy Data Bank (EMDB) with accession code EMD-0902. All other data relating to this study are available from the corresponding authors on reasonable request.

41. Chun, E. et al. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20**, 967–976 (2012).
42. Roth, C. B., Hanson, M. A. & Stevens, R. C. Stabilization of the human β_2 -adrenergic receptor TM4–TM3–TM5 helix interface by mutagenesis of Glu122³⁴¹, a critical residue in GPCR structure. *J. Mol. Biol.* **376**, 1305–1319 (2008).
43. Lv, X. et al. In vitro expression and analysis of the 826 human G protein-coupled receptors. *Protein Cell* **7**, 325–337 (2016).
44. Cherezov, V. et al. Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 μ m size X-ray synchrotron beam. *J. R. Soc. Interface* **6**, S587–S597 (2009).
45. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
46. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
47. Koehler Leman, J., Mueller, B. K. & Gray, J. J. Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics* **33**, 754–756 (2017).
48. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
49. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
51. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
52. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
53. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
54. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
55. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
56. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
58. Zhang, S. et al. Molecular matchmaking between the popular weight-loss herb *Hoodia gordonii* and GPR119, a potential drug target for metabolic disorder. *Proc. Natl Acad. Sci. USA* **111**, 14571–14576 (2014).

59. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
60. Yuan, Y., Pei, J. & Lai, L. LigBuilder 2: a practical de novo drug design approach. *J. Chem. Inf. Model.* **51**, 1083–1091 (2011).
61. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
62. The PyMOL Molecular Graphics system v1.8.x (Schrödinger, 2015).

Acknowledgements This work was supported by the National Key Research and Development Program of China, grant 2018YFA0507000 (to F.X. and S.Z.), National Natural Science Foundation of China (31525007 to M.L., 81861128023 to F.X. and 31971178 to S.Z.), Shanghai Municipal Education Commission–Gaofeng Clinical Medicine Grant Support (20181711 to J.W.) and Shanghai Outstanding Academic Leader funding (19XD1422800 to F.X.). The diffraction data were collected at BL41XU at SPring-8 with JASRI proposal 2019A2704. We thank J. Liu, N. Chen and L. Xue of the BV facility at the iHuman Institute, ShanghaiTech University for support with protein expression; the staff of the Electron Microscopy System and Mass Spectrometry System at Shanghai Institute of Precision Medicine for technical support and assistance in data collection; M. Cao for help with the collection and analysis of cryo-EM data; Q. Sun and the Bio-EM facility at ShanghaiTech University for technical support; V. Katritch and P. Popov for suggestions on construct mutations; M. Hanson for help with processing of X-ray data; and R. C. Stevens for encouraging this work.

Author contributions X.L. performed cloning, protein purification, crystallization, X-ray data collection, mutagenesis, cAMP functional assays and structural analysis; M. Li, N.W. and X.L. purified the GPR52 and mini-G_s–Nb35 proteins; M. Li prepared cryo-EM samples, collected data and determined structures; Y.W. performed structural analysis, molecular docking and structural similarity network analysis; J.W. performed model building and refinement; Z.L. processed the diffraction data with HKL2000 and solved the three crystal structures; S.G. designed ECL2 peptides and assisted with the data processing from cell-based functional assays; G.-W.H. was responsible for structure quality control; S.L. assisted with cryo-EM data collection and analysis; Y.Y. assisted with X-ray data collection; X.W. assisted with cell-based functional assays; X.X. supervised functional assays; Y.C. supervised structural analysis; S.Z. supervised structural analysis, molecular docking and structural similarity network analysis; and F.X. conceived the project and designed and supervised all experiments. All authors contributed to data interpretation and preparation of the manuscript. X.L., F.X. and M. Lei wrote the manuscript and F.X., M. Lei and J.W. orchestrated the project.

Competing interests The authors declare no competing interests.

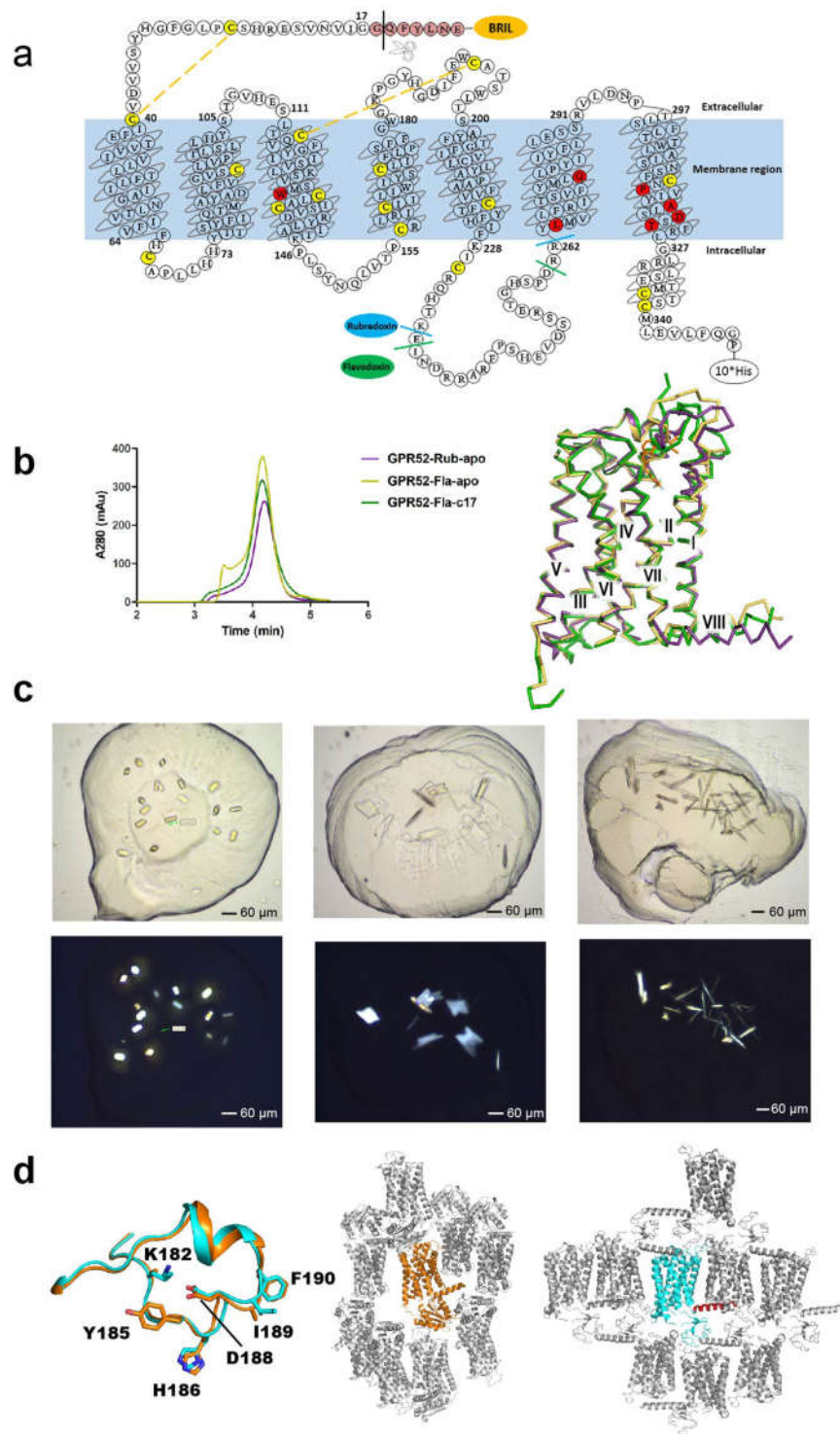
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2019-0>.

Correspondence and requests for materials should be addressed to J.W., M.L. or F.X.

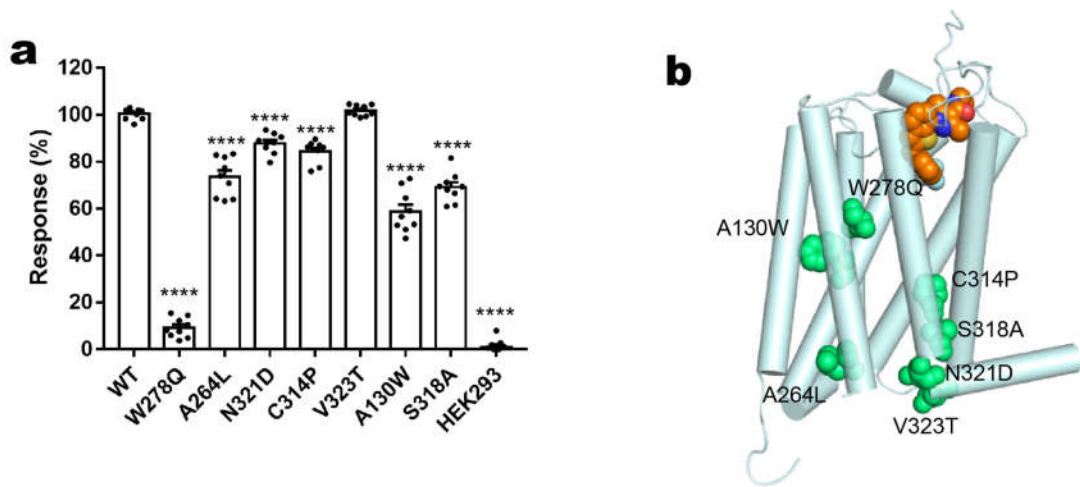
Peer review information *Nature* thanks Bryan Roth and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Engineering and crystallization of GPR52. **a**, Schematic of the GPR52 constructs that were used for crystallization (residues 17–340). Thermostabilizing mutations (red) are A130W, A264L, W278Q, C314P, S318A, N321D and V323T. Cysteine residues (yellow), TEV cleavage site (pink) and disulfide bonds (orange dashed lines) are shown. **b**, Left, analytical size-exclusion chromatography of GPR52. Experiments were repeated three times with similar results. Right, superposition of GPR52-Fla-apo, GPR52-Rub-apo

and GPR52-c17 structures shows the overall conserved helical arrangement. **c**, Crystal images of the GPR52-c17 complex (left), GPR52-Rub-apo (middle) and GPR52-Fla-apo (right). Experiments were repeated three times with similar results. **d**, ECL2 comparison of two apo structures (GPR52-Fla-apo, orange; GPR52-Rub-apo, cyan). The crystal packing of GPR52-Fla-apo (middle) and GPR52-Rub-apo (right) is also shown. Helix 8 of GPR52-Rub-apo is highlighted in red.



c Functional potency (pEC₅₀)

	WT	A130W	A264L	W278Q	C314P	S318A	N321D	V323T
c17	7.51±0.23	6.86±0.15	7.29±0.10	6.82±0.25	6.73±0.09	7.39±0.16	7.11±0.15	7.98±0.20

	ICL3-Rub	ICL3-Fla	GPR52-Rub-apo	GPR52-Fla-apo	Cryo-EM construct	K182E
c17	NA	NA	NA	NA	6.59±0.09	6.19±0.19

	C193A	△ECL2(182-198)	△182-190(GS)	△191-199(GS)	HEK293T
c17	7.55±0.15	NA	NA	6.08±0.05	NA

d Relative surface expression (%)

	WT	A130W	A264L	W278Q	C314P	S318A	N321D	V323T
Expression	100±12.9	105.9±12.4	77.3±5.5	87.3±1.4	100±20.5	88.6±17.7	72.3±8.5	54.5±10.9

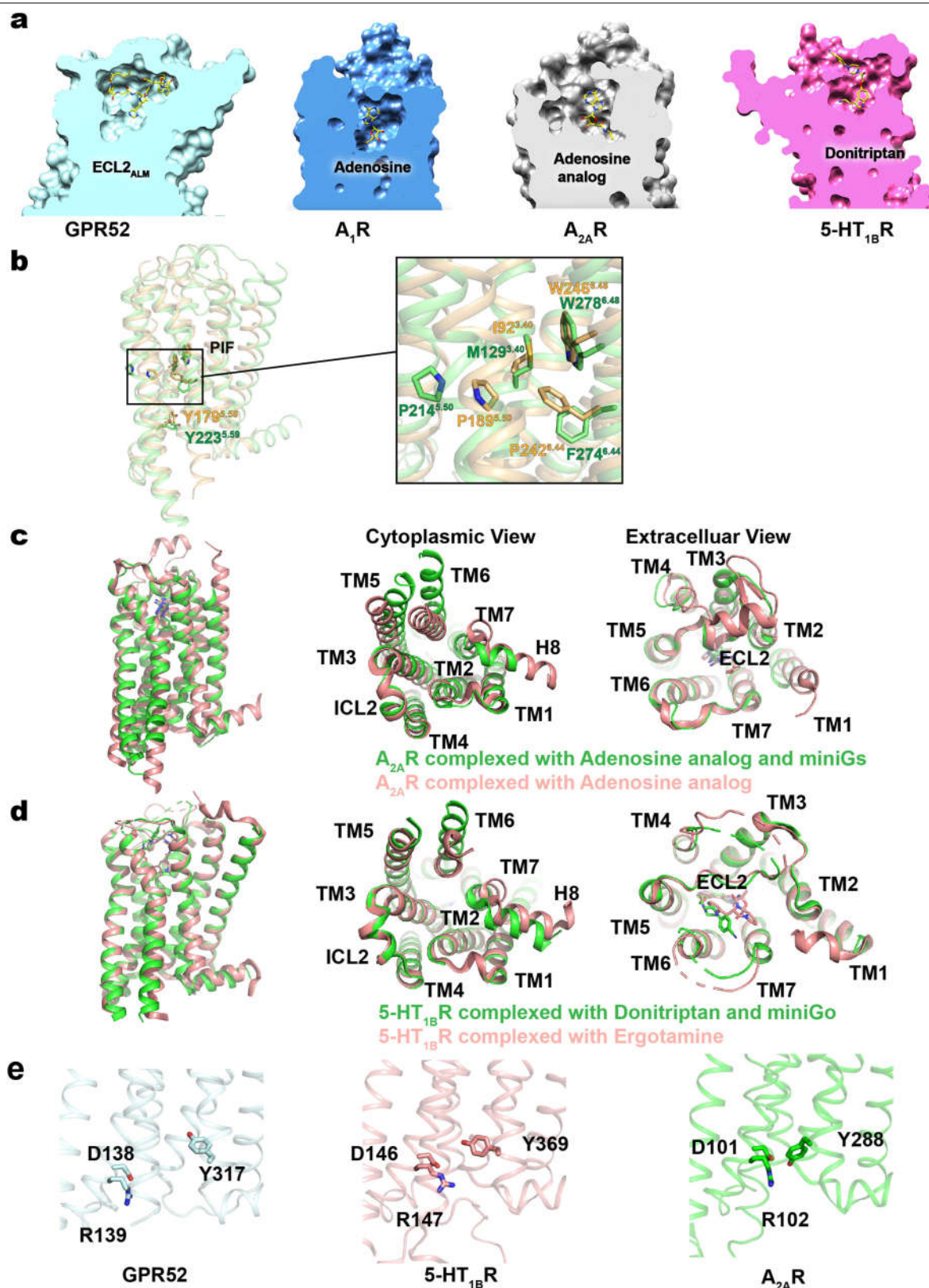
	ICL3-Rub	ICL3-Fla	GPR52-Rub-apo	GPR52-Fla-apo	Cryo-EM construct	K182E
Expression	66.8±6.2	70.0±1.6	102.3±6.8	96.4±14.7	78.6±15.0	36.3±6.3

	C193A	△ECL2(182-198)	△182-190(GS)	△191-199(GS)	HEK293T
Expression	48.0±4.6	69.6±11.0	30.5±3.7	38.2±10.1	NA

Extended Data Fig. 2 | Effects of GPR52 mutations on the potency of c17.

a, Basal activity of GPR52 mutants. Response-level values were compared with wild-type GPR52 by two-way ANOVA without repeated measures, followed by Dunnett's post hoc test (*****P* < 0.0001). Data are mean ± s.e.m. (*n* = 3) **b**, Mapping of mutated residues (green) onto GPR52 crystal structures. c17 is

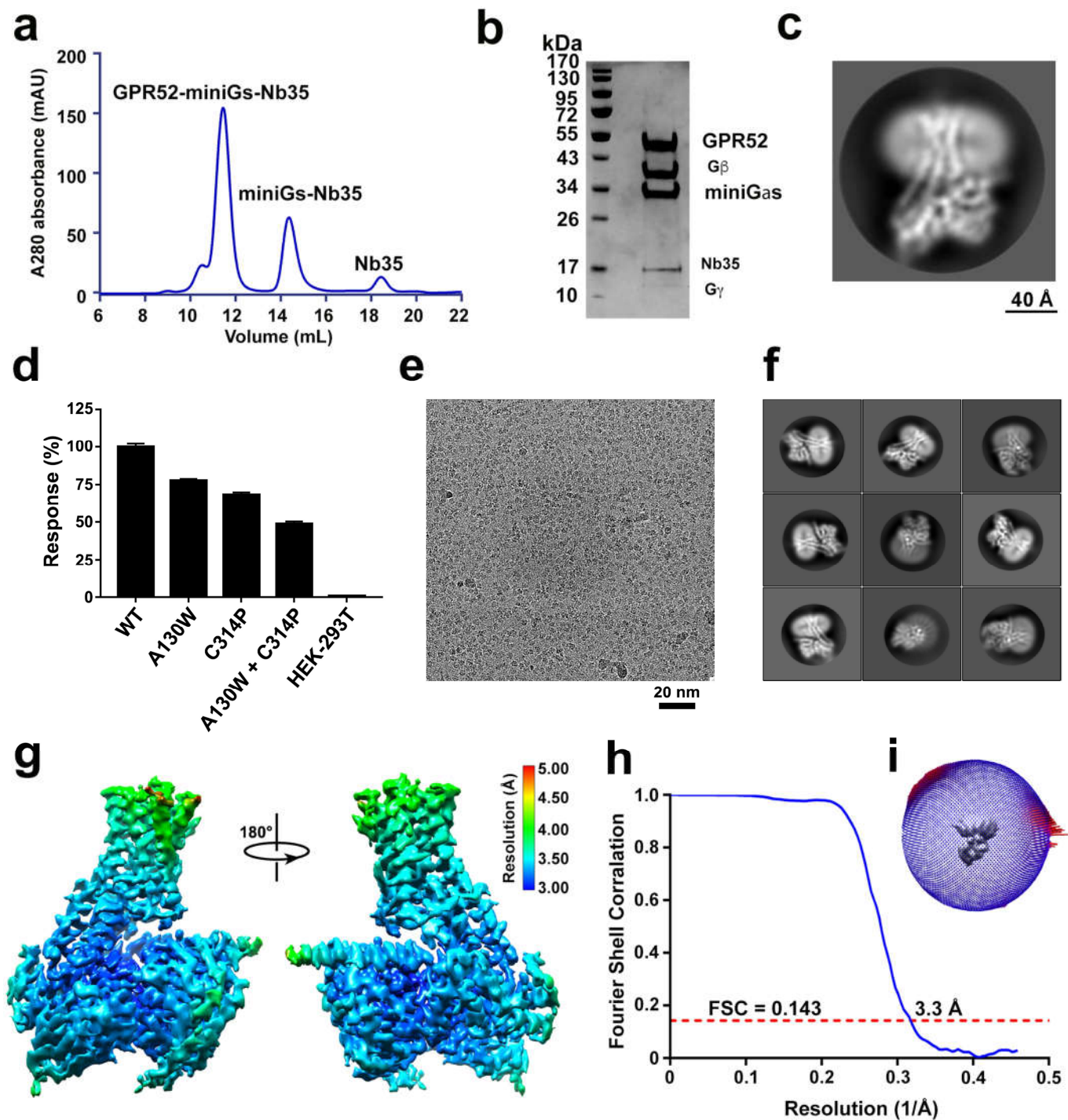
shown in orange. **c**, Summary of functional potency (pEC₅₀) values of c17 on the GPR52 mutants. Data are mean ± s.e.m. (*n* = 3). **d**, Relative surface expression levels of mutant constructs were monitored by a FACS staining assay (Methods) and normalized to the expression levels of wild-type GPR52. Data are mean ± s.e.m. (*n* = 3). NA, not available.



Extended Data Fig. 3 | Comparison of GPR52 with other class-A GPCRs.

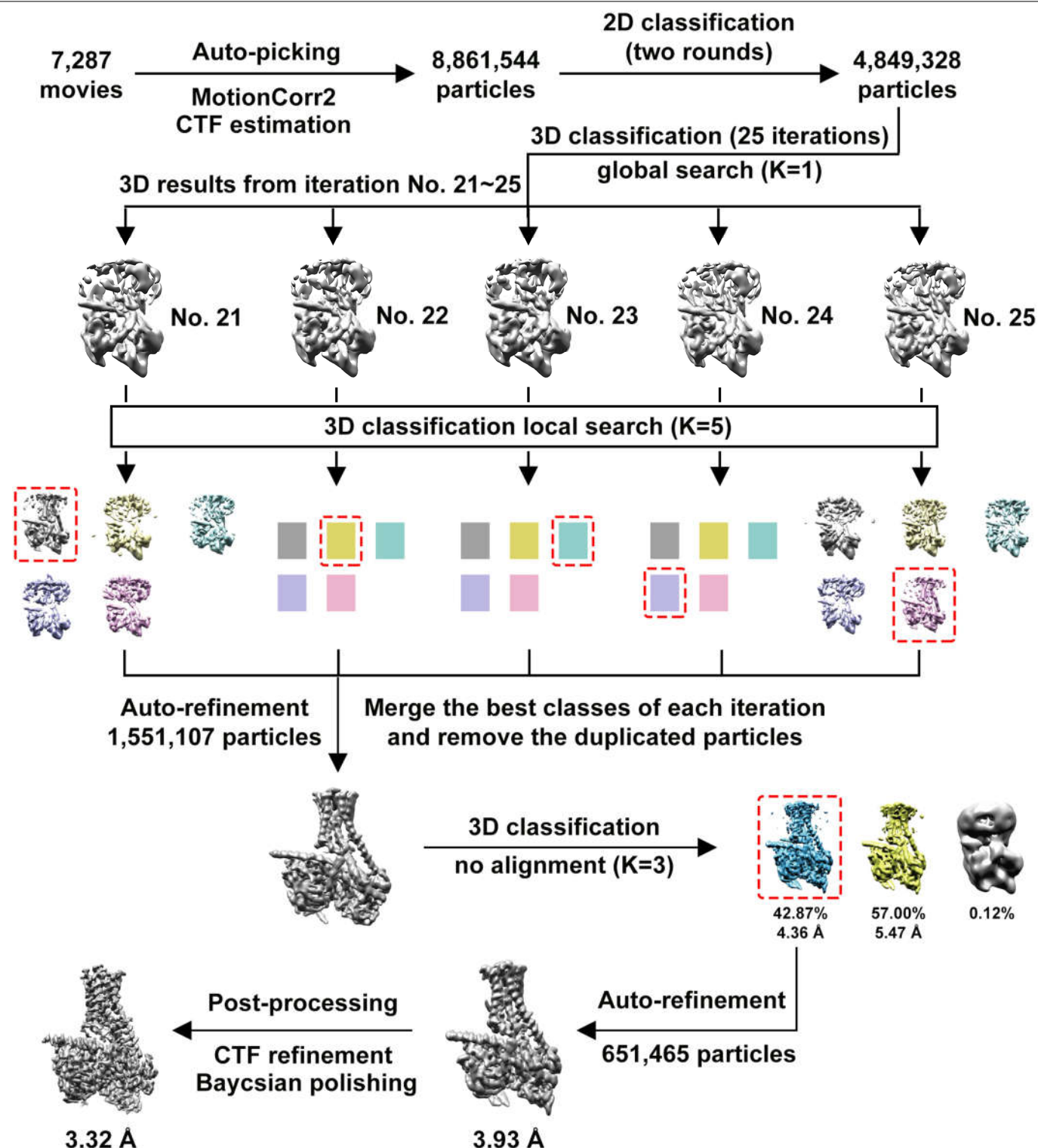
a, Comparison of the ECL2 ALM-occupied orthosteric binding pocket of GPR52 with agonist-bound pockets of A₁R (PDB 6D9H), A_{2A}R (PDB 6GDG) and 5-HT_{1B}R (PDB 6G79). **b**, PIF motif comparison of mini-G_s-coupled GPR52 (green) and A_{2A}R (yellow; PDB 6GDG). **c**, Side view (left), cytoplasmic view (middle) and extracellular view (right) of A_{2A}R in the mini-G_s-coupled state (green, PDB

6GDG) compared with the adenosine analogue (NECA; PDB 2YDV)-bound state (pink). **d**, Side view (left), cytoplasmic view (middle) and extracellular view (right) of 5-HT_{1B}R in the mini-G_o-coupled state (in complex with donitriptan) (green; PDB 6G79) compared with the ergotamine-bound state (pink; PDB 4IAR). **e**, DRY motif of GPR52-apo, 5-HT_{1B}R-ergotamine (PDB 4IAR) and A_{2A}R-NECA (PDB 2YDV).

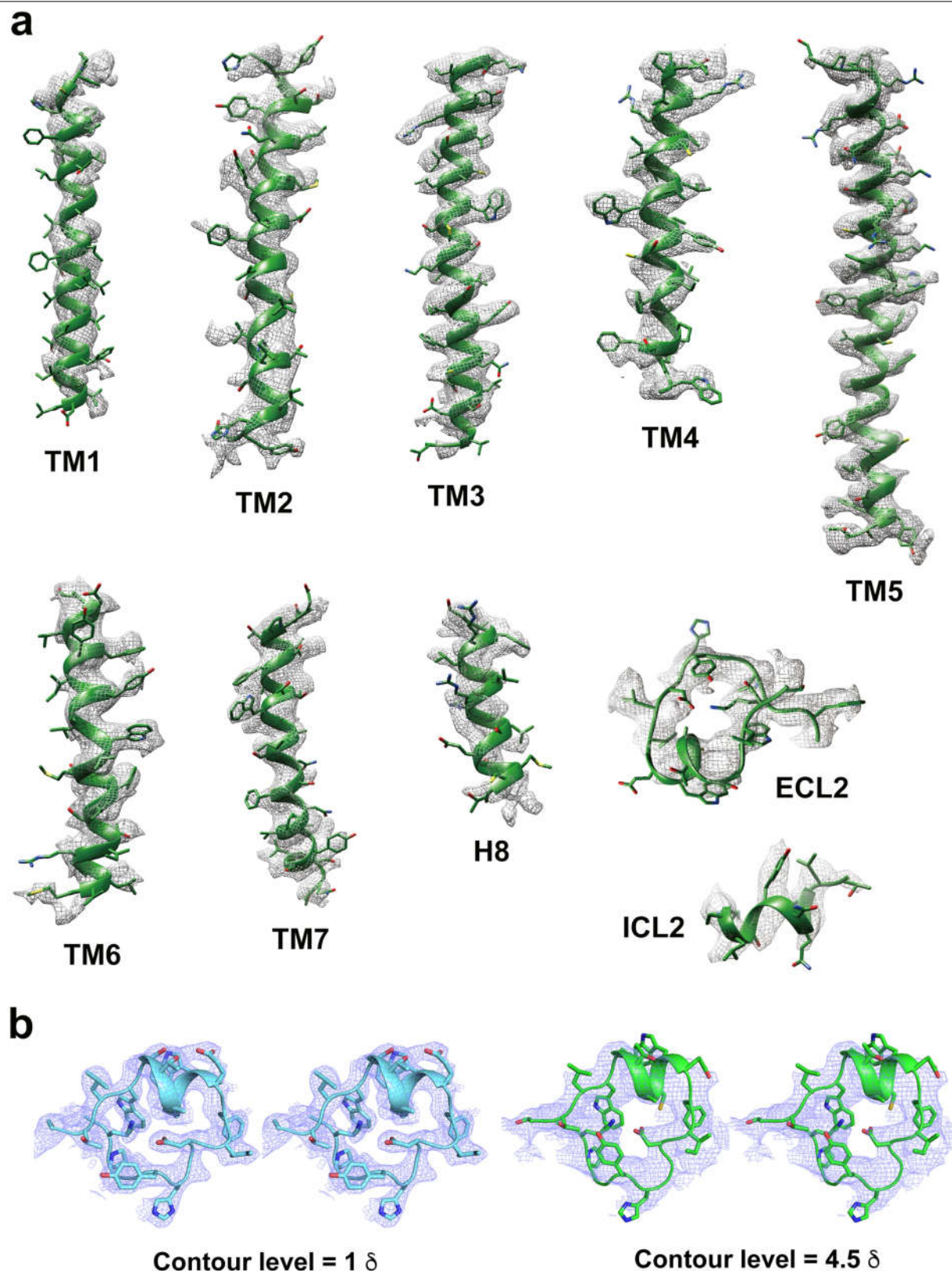


Extended Data Fig. 4 | Cryo-EM analysis of the GPR52-mini-G_s-Nb35 complex. **a, b**, Size-exclusion chromatography profile (**a**) and corresponding SDS-PAGE gel (**b**) of the purified GPR52-mini-G_s-Nb35 complex. Experiments were repeated three times with similar results. **c**, Representative reference-free 2D cryo-EM average of the GPR52-mini-G_s-Nb35 complex. **d**, GPR52 with point mutations A130W^{3,41} and C314P^{7,50} maintained around 50% of the activity relative to the wild-type protein, according to the cAMP response level. Data

are mean \pm s.e.m. ($n=3$). **e**, Representative cryo-EM micrograph of the GPR52-mini-G_s-Nb35 complex. **f**, Reference-free 2D averages of the GPR52-mini-G_s-Nb35 complex. **g**, Final 3D density map coloured according to the local resolution. **h**, Gold-standard FSC curves, showing the overall nominal resolution at 3.3 Å. **i**, Angular distribution of the particles used for the final reconstruction of the GPR52-mini-G_s-Nb35 complex.

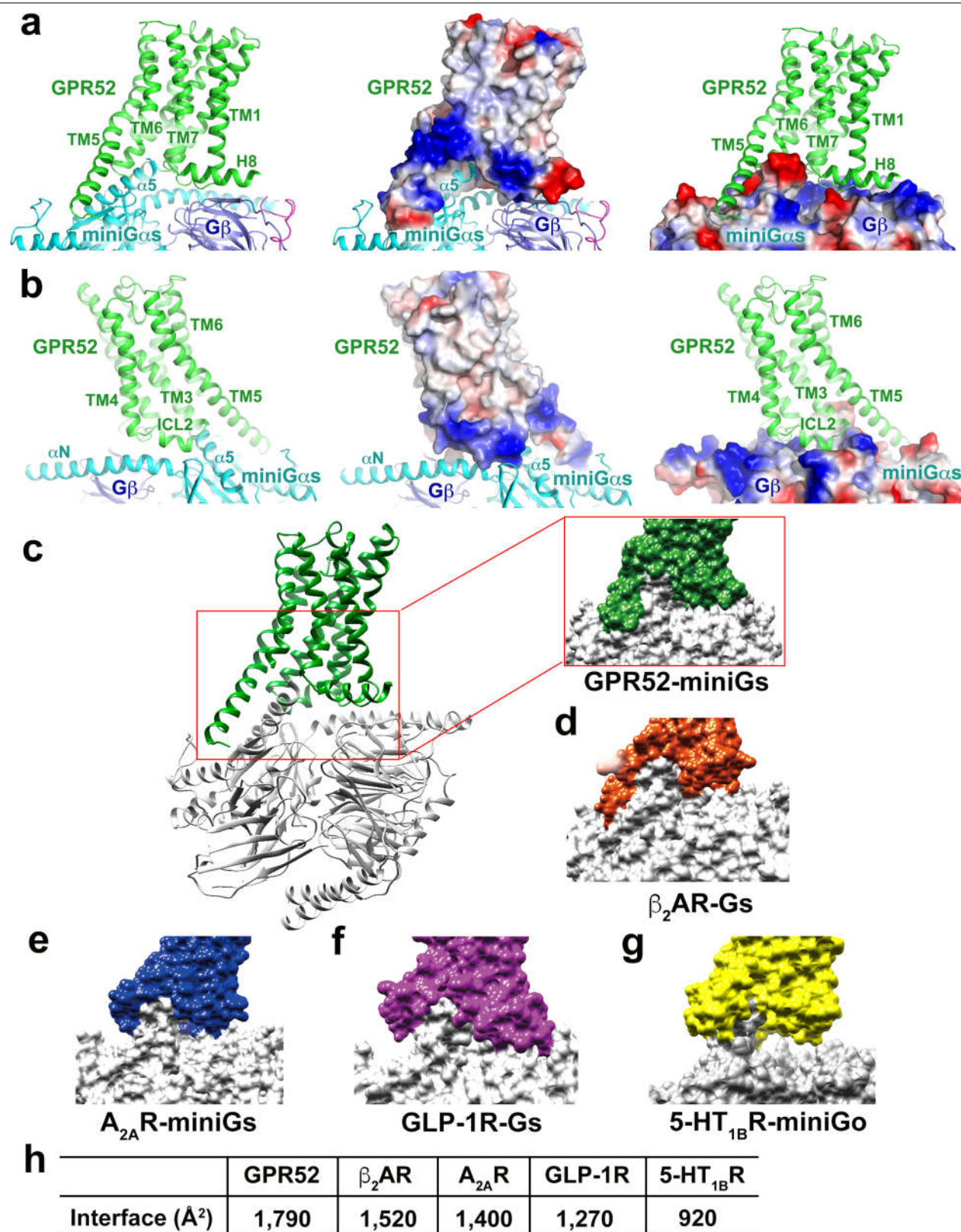


Extended Data Fig. 5 | Flow chart for the cryo-EM data processing and structure determination of the GPR52-mini-G_s-Nb35 complex. See Methods for details. The final reconstruction has an average resolution of 3.3 Å. All the images in this figure were created in UCSF Chimera.



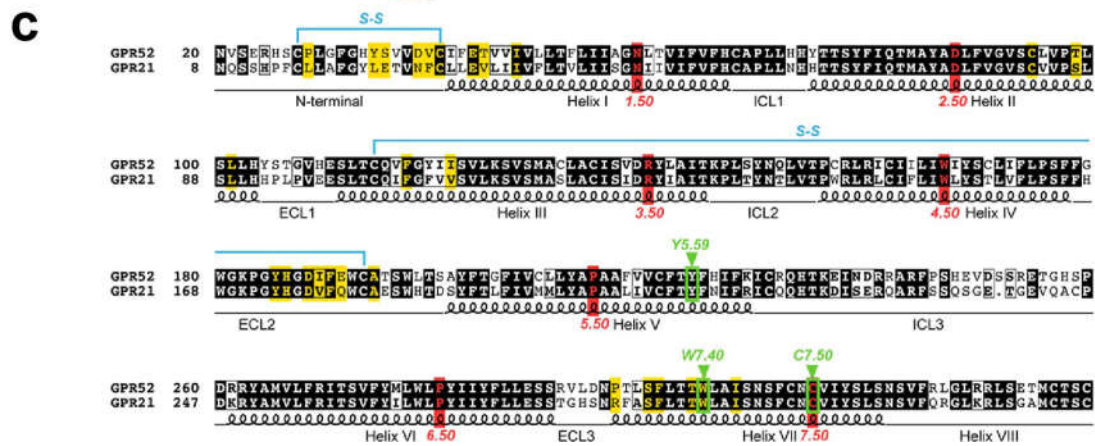
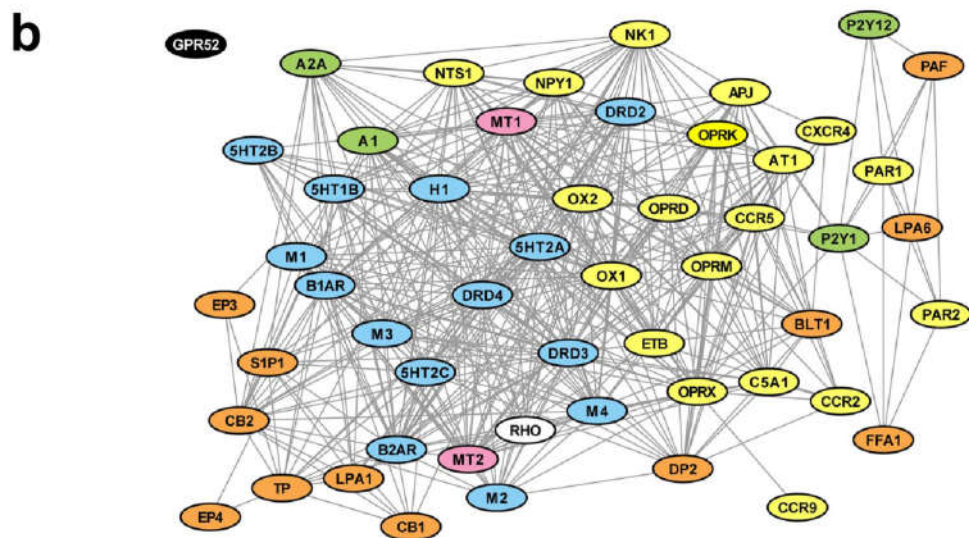
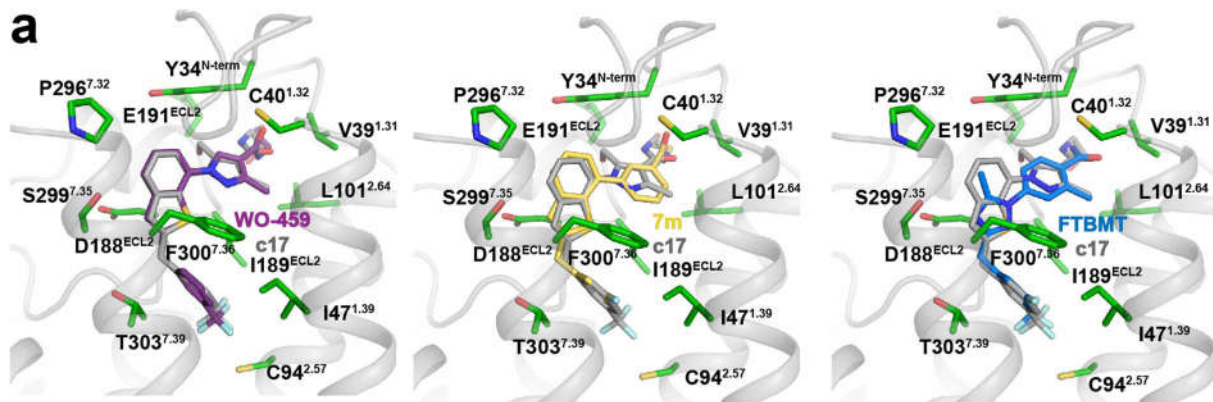
Extended Data Fig. 6 | Cryo-EM map quality and ECL2 comparison. **a**, Atomic model of GPR52 transmembrane helices, ECL2 and ICL2 in the cryo-EM density map. The molecular model is shown in stick representation and the cryo-EM

map as mesh. **b**, Stereo views of the electron density maps of ECL2. Left, the $2F_o - F_c$ map of ECL2 from the GPR52-Rub-*apo* crystal structure. Right, the electron density map of ECL2 from the GPR52-mini-G_s complex structure.



Extended Data Fig. 7 | Comparison of the GPR52–mini-G_s interface with that of other GPCR–G-protein complexes. **a, b, Front view (**a**) and back view (**b**) of the GPR52–mini-G_s interface. GPR52 (centre) and mini-G_s (right) are in surface representation and coloured according to the electrostatic potential (blue, positive; red, negative). **c**, The GPR52–mini-G_s interface. GPR52 and mini-G_s are**

in cartoon representation and coloured in green and grey, respectively. A magnified view of the interface is shown on the right in surface representation. **d–g, Magnified views of the interface between other receptors and G proteins in surface representation. **h**, Buried surface area of the interfaces between receptors and G proteins, calculated by PyMOL.**



Extended Data Fig. 8 | Docking position of GPR52 agonists, structural similarity network and sequence alignment of GPR52. a. Docking position of GPR52 agonists with different scaffolds: WO-459 (left); 7m (centre) and FTBMT (right). The PR52 residues that are involved in ligand binding are shown as

green sticks and c17 is shown as greysticks, for reference. **b**, Structural similarity network of class-A GPCRs with reported inactive structures. **c**, Sequence alignment of GPR52 and GPR21 (yellow, less than 5.0 Å to ligand; green indicates key residues for structural features).

Extended Data Table 1 | Data collection and structure refinement statistics

Structure	GPR52-Fla-apo ^a	GPR52-Rub-apo ^b	GPR52-Fla-cl7 ^c
PDB ID	6LI1	6LI2	6LI0
Data collection			
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>I</i> 222	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	66.65, 79.89, 148.30	77.22, 113.23, 138.62	59.97, 88.36, 156.28
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	28.41-2.90 (2.98-2.90)	29.25-2.80 (2.90-2.80)	41.89-2.20 (2.26-2.20)
<i>R</i> _{merge}	0.24 (2.75)	0.16 (0.97)	0.18 (2.89)
<i>I</i> / σ < <i>I</i> >	11.8 (1.0)	13.2 (1.7)	18.3 (1.0)
Completeness (%)	97.60 (98.80)	96.9 (81.10)	99.37 (96.11)
Redundancy	9.2 (9.4)	8.2 (8.0)	26.5 (17.5)
CC _{1/2}	0.99 (0.40)	1.02 (0.90)	0.99 (0.44)
Refinement			
Resolution (Å)	28.41-2.90	29.25-2.80	41.89-2.20
No. reflections	16,462	11,830	40,732
<i>R</i> _{work} / <i>R</i> _{free}	0.244 / 0.267	0.241 / 0.263	0.194 / 0.220
No. of atoms			
Protein	3,339	2761	3479
Ligand	n/a	n/a	34
Lipids and others	79	195	397
B-factors (Å ²)			
Wilson / Overall	97.0 / 88.1	76.8 / 68.2	60.3 / 88.5
Protein	88.5	67.4	80.4
Ligand	n/a	n/a	69.3
Lipids and others	71.3	80.2	106.6
R.M.S deviations			
Bond lengths (Å)	0.008	0.008	0.011
Bond angles (°)	1.54	1.44	1.66

^a20 crystals used for structure determination. Values in parentheses are for the highest-resolution shell.^b25 crystals used for structure determination. Values in parentheses are for the highest-resolution shell.^c57 crystals used for structure determination. Values in parentheses are for the highest-resolution shell.

Extended Data Table 2 | Cryo-EM data collection and refinement statistics

	GPR52-miniGsβγ-Nb35 (EMDB-0902) (PDB 6LI3)
Data collection and processing	
Magnification	75,000
Voltage (kV)	300
Electron exposure (e-/Å²)	40.0
Defocus range (µm)	-1.0 to -2.6
Pixel size (Å)	1.09
Symmetry imposed	C1
Initial particle images (no.)	8,861,544
Final particle images (no.)	651,465
Map resolution (Å)	3.3
FSC threshold	0.143
Map resolution range (Å)	3.0-7.7
Refinement	
Initial model used (PDB code)	6GDG
Model resolution (Å)	4.1
FSC threshold	0.143
Model resolution range (Å)	n/a
Map sharpening B factor (Å²)	-121
Model composition	
Non-hydrogen atoms	8,289
Protein residues	1,045
Ligands	0
B factors (Å²)	
Protein	94.58
Ligand	n/a
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.543
Validation	
MolProbity score	1.50
Clashscore	7.53
Poor rotamers (%)	0.00
Ramachandran plot	
Favored (%)	97.58
Allowed (%)	2.42
Disallowed (%)	0.00

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection EPU-v2.2.0

Data analysis HKL2000, Phaser, Global phasing, Phenix-v1.16, Refmac5, Buster, COOT-v0.8.9.2, MolProbity, PyMOL-v1.8.x, UCSF Chimera, GROMACS, Schrödinger Suite-v2019-2, Cavity-v1.1, Cytoscape-v3.7.1, Esprito-v3.0, Gctf_v1.06, MotionCor2-v1.0.0, RELION-v3.0.b2, Rosetta, Prism 7.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coordinates and structure factors for GPR52-Fla-c17, GPR52-Fla-apo, GPR52-Rub-apo and GPR52-miniGs-Nb35 have been deposited in the Protein Data Bank (PDB) with the accession number 6LI0, 6LI1, 6LI2, 6LI3 respectively. The cryo-EM 3D maps of the GPR52-miniGs-Nb35 complex was deposited in EMDB database with accession code EMD-0902. All other data relating to this study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For all the functional assay, three independent experiments (n=3) were carried out.
Data exclusions	No data were excluded
Replication	Experimental findings were reliably reproduced
Randomization	Randomization was not required
Blinding	Blinding was not performed

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	sf9 cell line, Invitrogen, Cat# 11496-015; HEK293 cells also from Invitrogen.
Authentication	No authentication required
Mycoplasma contamination	All cell lines tested are negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Author Correction: Quantifying secondary transport at single- molecule resolution

<https://doi.org/10.1038/s41586-020-2029-y>

Correction to: *Nature* <https://doi.org/10.1038/s41586-019-1747-5>

Published online 13 November 2019



Check for updates

**Gabriel A. Fitzgerald, Daniel S. Terry, Audrey L. Warren,
Matthias Quick, Jonathan A. Javitch & Scott C. Blanchard**

In this Article, the LIV-BP sensors used were site-specifically labelled with self-healing LD555P and LD655 organic fluorophores, not LD550 and LD650 as indicated. Also, we site-specifically labelled LIV-BP mutant D181C, not E181C. The sentence “The rate of single-turnover transport decreased by nearly an order of magnitude when the external pH was **increased from 6 to 8** (Fig. 3c)” should read “The rate of single-turnover transport decreased by nearly an order of magnitude when the external pH was **decreased from 8 to 6** (Fig. 3c)”. These errors have no effect on the results presented or the interpretations and conclusions made. These errors have been corrected online.



ANNA SCHROLL/FOTOGRAFIA/IG/GETTY

Research using highly pathogenic organisms often needs to be done in full protective suiting.

BEHIND THE SCENES IN A BIOSAFETY OFFICE

It's never a dull day for those tasked with keeping biological research safe for all. **By Kendall Powell**

David Gillum received a flood of phone calls in January, after his university announced that someone in the campus community had been diagnosed with the COVID-19 coronavirus.

Gillum is the director of environmental health and safety at Arizona State University in Tempe, and many of the calls were from

biosafety officers around the globe. They were seeking his expertise and guidance in the event that their own institutions were in the same situation.

"Emergencies happen, and it's up to you to be the one to keep people calm," says Gillum. But most days in the life of an institutional biosafety officer are much less dramatic.

Generally, the officers collaborate with their institute's scientists to work out how to safely conduct biological research, such as that involving infectious agents or modified DNA. But their roles have expanded over the past decade, as biological research and its hazards have become more complex. Large research institutes and campuses can have dozens of

officers; smaller institutes often cover the role in their safety offices.

In Mexico and other Latin American countries, the position is often given the title of biosafety responsible, says Luis Ochoa Carrera, head of the Epidemiological Surveillance and Research Laboratory Network at the Mexican Institute for Social Security in Mexico City.

“As a graduate student, I thought biosafety was that person with their clipboard to tell you what you were doing wrong,” says Meghan Seltzer, manager of safety, health and security at the Howard Hughes Medical Institute’s Janelia Research Campus in Ashburn, Virginia. “But there is so much more to it.” In a typical week, an officer might write a risk assessment for a project that would add chemically synthesized designer molecules to living cells; review the design and construction of tissue-culture room suitable for work that requires strict controls; or assess the risks to society from research that modifies the influenza virus. “There’s never a dull moment in our jobs,” says Danielle Rintala, biological-safety officer for the University of Wisconsin–Milwaukee. “There’s always something to learn about.”

Many paths

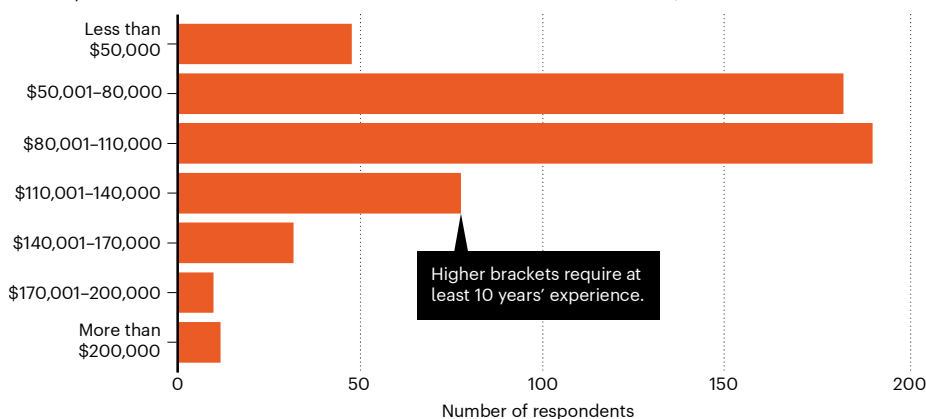
Few officers hold degrees relating specifically to biosafety, and in the United States, at least, most qualifications tend to be offered by institutes other than universities. There are, however, many courses that lead to accreditation as a biosafety professional, including the National Biosafety and Biocontainment Training Program at the US National Institutes of Health (NIH) in Bethesda, Maryland, and a programme offered by the Biosafety Training Institute at the University of Edinburgh, UK. Seltzer, for instance, completed the two-year NIH programme.

Biosafety officers can have backgrounds in life sciences, public health, medicine, engineering, education and emergency response, among others, says Seltzer, who is also a member of a careers task force for ABSA International, the Association for Biosafety and Biosecurity in Mundelein, Illinois. They tend to hold a range of qualifications, from undergraduate diplomas through to doctorates. Those in the job recommend at least a master’s-level knowledge of microbiology or molecular biology – because those fields cover the most risky research. That level of education is important, they say, because the training will be used daily in critically evaluating complex scientific ideas, anticipating potential hazards and executing research protocols. Ochoa Carrera, for example, trained in chemistry, pharmacy and biology, and has a master’s in public-health management. Also important, he and others say, is laboratory experience working with infectious fungi, bacteria, viruses or parasites.

And Toshinori Tanaka, biosafety specialist for the Okinawa Institute for Science of

COMFORT ZONE

Biosafety officers have standard work weeks with set hours. Most US officers earn \$50,000–110,000.



Technology Graduate University (OIST) in Japan, has a PhD in plant science. He says that his posts working for a local government office on forestry and soil science and on agricultural policy in Nagasaki gave him insights into government regulations and priorities.

Competitive salaries

In 2013, Gillum and his colleagues found that salaries for biosafety professionals working in the United States ranged from up to US\$70,000 per year for entry-level positions to around \$110,000 for those with 15 years or more experience (see ‘Comfort zone’, and D. Gillum *et al. Appl. Biosaf.* **18**, 106–115; 2013). In Austria, early-career professionals can expect to earn a salary of €69,040 (US\$75,600) – similar to that of a postdoctoral fellow, says Gabriel Ó Riordáin, head of scientific support for the Austrian Academy of Science’s Research Center for Molecular Medicine in Vienna.

SAFETY TRAITS FIRST

The mindsets and skills of great biosafety professionals.

- Willing to wear personal protective equipment. Officers often set the tone for safety culture at their institutions.
- Enjoys learning. Biosafety professionals analyse research and protocols across the entire realm of biology, especially leading-edge technologies.
- Listens closely. Effective professionals don’t pass judgement on mistakes, but rather offer reassurance and help researchers learn from them.
- Mechanically inclined. Officers might need to assess, and possibly repair, lab equipment ranging from airflow vents to pipettors.
- Cool under pressure. Biosafety officers must be able to remain calm in a crisis, and to think clearly and quickly to solve a problem.

Zachary Wilson, assistant biosafety officer at the Anschutz Medical Campus of the University of Colorado Denver, appreciates the fact that his salary is not tied to grant funding, yet he is still involved in analysing and shaping research projects. He says he gets to learn what all the biologists on campus are investigating, ranging from what mechanisms the tuberculosis bacterium uses to survive antibiotic treatment to how autoimmune disorders develop in mice with humanized immune systems. “I didn’t want to study one protein in one bacterium for my entire life,” he says.

Many biosafety officers rank their involvement in research, albeit in a supporting role, as one of the key attractions of the job. “One of the things I’ve loved most about this job is that I’m still involved in and helping the research community,” says Andrea Ladd, assistant director of the environment, health and safety office at the University of Wisconsin–Madison.

Rintala, too, enjoys learning about all the biology research across her campus. “I see every single person working with biological materials,” she says, from engineers developing microscopy techniques to researchers studying pollution in the nearby Great Lakes.

Peter Farina, director of safety at National Jewish Health hospital and research institute in Denver, finds that his job sometimes crosses over into the clinical realm. One of the institution’s specialties is respiratory diseases, and many researchers work closely with the tuberculosis bacterium. He often needs to work with the hospital’s infection-prevention officer to make sure that protections are in place for researchers, clinicians and patients, and that wards and clinics are disinfected properly after procedures.

Days spiced with variety

Many researchers find the position appealing for its standard working week, with set hours and flexibility to work from home on some days. Before getting her current role, Ladd had spent 11 years as a principal investigator studying heart development and disease at

the Cleveland Clinic's Lerner Research Institute in Ohio. She was spending 40 hours per week writing grants and another 30 hours at the bench. It was an unsustainable schedule and she knew it had to change. Even though she had no experience as a biosafety officer or any training in microbiology, she was aware of the importance of biosafety procedures and was comfortable with the administrative side of research and how to manage teams and budgets. So she took a gamble and applied for her job at the University of Wisconsin–Madison. “I pitched hard in the interview that I could learn any of the science I needed to learn,” she says.

Many biosafety officers work alongside colleagues who oversee the handling and disposal of hazardous waste, monitoring of air and water supplies, handling of chemicals and radiation and employee health and well-being. They are also members of committees that review and approve research projects involving biological materials such as infectious agents, recombinant DNA or synthetic nucleic acids, as well as committees that oversee the care of research animals.

Reviewing research protocols helps officers to determine the biosafety level (BSL) at which a research team should be working, ranging from 1 for the least hazardous to 4 for the most. That means assessing which experiments must be done in a controlled cabinet, what protective gear should be worn and how biohazardous waste should be disposed of.

Another large part of the job involves building relationships with laboratories on campus. “If I take the time to sit and learn why researchers want to do a particular experiment, it really helps me do my job,” says Seltzer. “It helps me solve the safety problem better, which helps them get their research done, and I get excited about it.”

Building trust is also crucial for times when something goes wrong. The most common incidents are needle pricks, eye splashes, cuts from sharp instruments, bites from research animals and spills. It's very rare for a biosafety officer to have to suit up in protective coveralls and respirators for an emergency. Like firefighters, however, they need to be prepared for such scenarios. And unusual incidents sometimes require officers to quickly find people on campus who can advise them on the science, the facility or both, says Farina.

That said, it doesn't hurt to be an “addict to adrenaline”, says Ochoa Carrera, who formerly worked as a BSL-3 coordinator for Mexico's Ministry of Health. “I absolutely like that part of the job. You need to be aware of the potential things that can happen inside or outside of the lab and be ready to respond to all types of threats.” His mettle was tested in 2013, when he led a mobile laboratory team to investigate a cholera outbreak in the rural region of Hidalgo for two months. The situation called for both his technical expertise in handling



David Gillum is a biosafety officer at Arizona State University in Tempe.

pathogens and his leadership skills. “You have to know how to handle the pressure,” he says.

Most of the job is about prevention. “It's always been our office's practice that we don't want to end up on the local news,” Rintala says. And officers lead annual training for lab workers, including when and how to use safety glasses and other protective gear, how to transport or pipette biological hazards, and

“One of the things I've loved most about this job is that I'm still involved in and helping the research community.”

how to use eye-washing stations and fire extinguishers. They also inspect equipment and procedures. Tanaka says that he is continually reminding researchers in tropical Okinawa not to wear open sandals in the lab.

But when incidents do occur, biosafety officers must remain calm, listen carefully and without judgement to what happened, and see to the researcher's immediate welfare (see ‘Safety traits first’). Rintala says that it's helpful to maintain a reassuring, non-critical demeanour and make sure that researchers don't feel blamed or shamed for making a mistake. “It's really important not to treat it as punishment,” she says, “but rather, to treat it as a learning experience.” That means that

researchers will be more likely to report problems and ask biosafety officers for help the next time. Once the situation is under control, biosafety officers need to investigate what went wrong and discuss how it could be prevented in the future.

Challenges and rewards

Tanaka says that one of the most rewarding and challenging parts of the job is deciding how to work safely with new technologies. “The more advanced the research is, the fewer regulations exist to cover it,” he says. In one such example, OIST researchers created a hybrid virus, combining parts from one that causes an animal disease and from another that is harmless. No regulations existed for handling such a hybrid. “The biosafety officer on site needs to [balance] safety measures in a way that also ensures the freedom of research.”

Biosafety officers describe the job as a great career option for researchers needing a change of pace but wanting to stay involved in research. Although it requires pivoting to a supporting role, they find that role very satisfying. And sometimes, they find themselves in an adrenaline-filled moment, suited up in protective gear.

“It might seem like a scene out of the movie *Contagion*,” Wilson says. “But if we don't feel comfortable working in those labs, then we're doing something wrong.”

Kendall Powell is a freelance writer in Lafayette, Colorado.

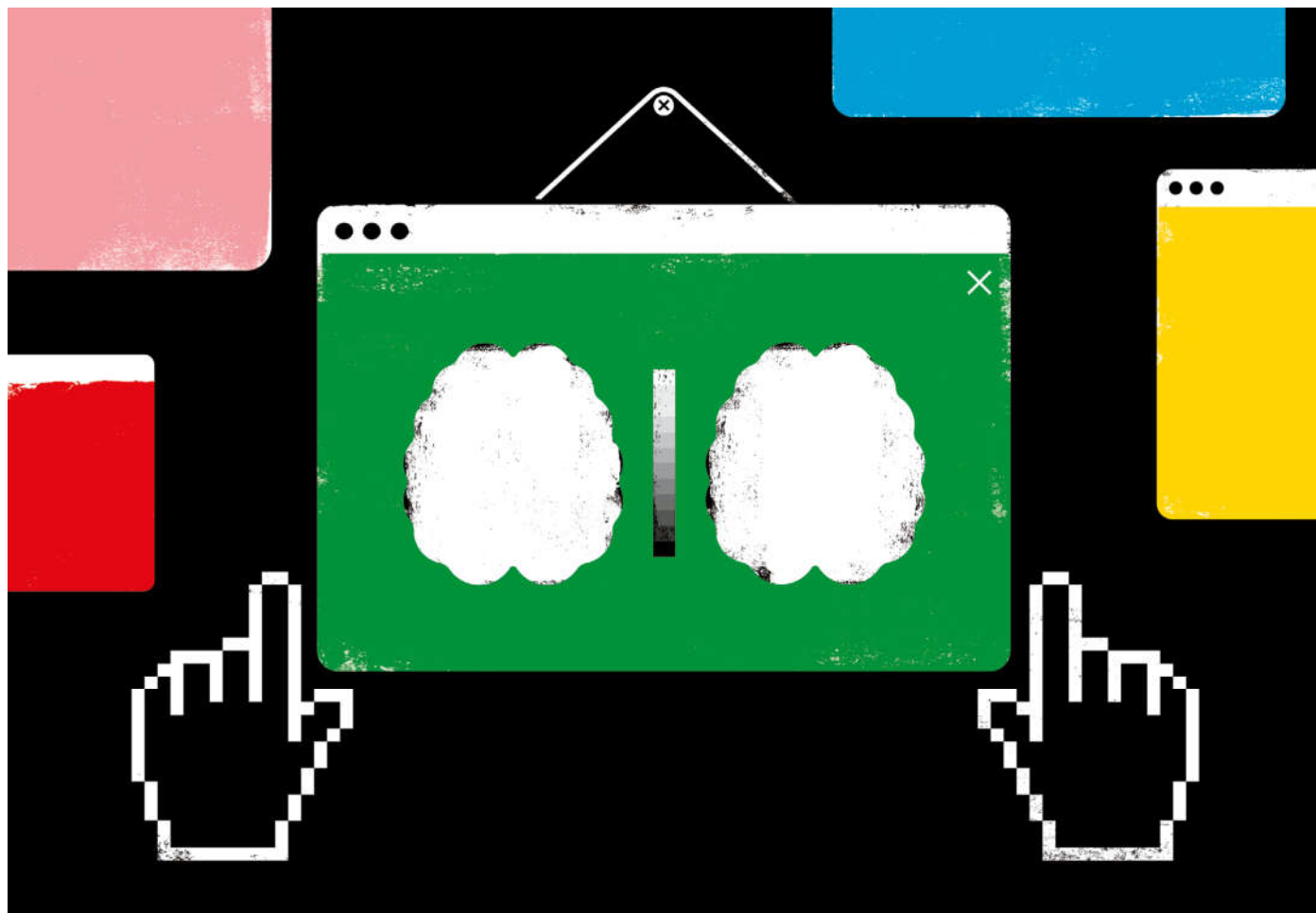


ILLUSTRATION BY THE PROJECT TWINS

A HOME FOR EVERY IMAGING DATA SET

Repositories let researchers store, share and access life-science images – and maybe even extract new findings. **By Amber Dance**

When Sjors Scheres set out to develop a tool to reverse flaws in cryo-electron microscopy images, he needed lots of data on which to test it. So Scheres, a structural biologist at the MRC Laboratory of Molecular Biology (LMB) in Cambridge, UK, turned to the Electron Microscopy Public Image Archive (EMPIAR), a database of raw images. There he downloaded, for free, data collected by the lab of Gabriel Lander, a structural biologist at Scripps Research in La Jolla, California.

Using his new technique, Scheres was able to squeeze sharper images from those data, improving¹ the resolution of one structure from 3.1 ångströms to 2.3 ångströms.

“That’s precisely why we posted the data,” says Lander. “We knew some brilliant people out there would be able to improve on

our processing.”

Services such as EMPIAR give researchers a central location in which to store, share and access a rapidly expanding corpus of biological images. “The data aren’t just one picture any more,” says Joshua Vogelstein, a neuro-statistician at Johns Hopkins University in Baltimore, Maryland. Movies, 3D images and microscope-based screening data can take up gigabytes or terabytes of storage, and can’t be e-mailed back and forth in the same way as individual TIFF or JPEG files. Moreover, grant agencies and journals increasingly require scientists to make their data available to all, but don’t necessarily offer to host them. EMPIAR and its kin fill that gap, and often provide a digital object identifier or other citation so researchers can get credit for their data.

“Are you struggling to load your images?”

asks Forrest Collman, a neuroscientist at the Allen Institute for Brain Science in Seattle, Washington. “Are you particularly struggling to share?” If so, he says, “looking into this kind of service makes sense for you”.

In 2019, when Collman spotted an odd-looking neuron in one of his electron-microscopy data sets², it was easy for him to send a colleague a link to that spot in the data repository, rather than a bulky file. She noticed another unique feature, and Collman identified a few similar cells. They might turn out to be a new type of neuron, Collman says.

There are a number of other image warehouses available, among them the Image Data Resource (IDR). Both it and EMPIAR are hosted by the European Molecular Biology Laboratory’s European Bioinformatics Institute (EMBL-EBI) in Hinxton, UK. Further options include,

but are not limited to, NeuroData, a platform that Vogelstein set up to host neuroanatomy files, and the Systems Science of Biological Dynamics (SSBD) database at Japan's RIKEN network of research institutes. Advocates expect these platforms to follow the model of established DNA- and protein-sequence resources such as GenBank and the Protein Data Bank, which have powered an array of analyses and spawned the field of bioinformatics.

"We're very early days," says Jason Swedlow, a quantitative cell biologist at the University of Dundee, UK. But he expects big benefits, both for scientists who download large image sets to feed data-hungry machine-learning algorithms and for those who might make new discoveries in others' data.

Share and share alike

It was a data-hungry scientific community that drove Kate McDole, a developmental biologist at the LMB, to use an image database.

McDole, then working at the Howard Hughes Medical Institute's Janelia Research Campus in Ashburn, Virginia, had imaged mouse embryos every five minutes as they developed, yielding terabytes of data and a high-resolution developmental atlas³ that has generated significant interest. "People are forever asking me, did you look at this tissue, did you look at that tissue?" So she looked for a way to share all those terabytes.

The journal offered only gigabytes of space, much less than McDole needed. ("Oh, gigabytes," she scoffs, "gigabytes are cute.") So she uploaded the atlas to the IDR, a free service developed by Swedlow and his colleagues. The data transfer took the better part of a week, she says. But now, anyone with a web browser can scroll through her data set, find their favourite tissues, or compare their results with hers. McDole herself often uses the IDR at conferences, to show colleagues data she doesn't carry on her laptop.

Such databases offer more than a storage location, says Jan Ellenberg, a cell and molecular biologist at EMBL in Heidelberg, Germany, and researchers shouldn't simply drop their data sets into small, project-specific archives or generic cloud storage. "Just dumping the data somewhere doesn't mean people can use it," Ellenberg explains. "You need to organize the data, you need to annotate it, and curate it." Browsers of McDole's data set, for instance, can scan the metadata to find out information such as the strain of mice she used and the specific fluorescent labels she imaged.

Patrick Combes, global technical leader for health care and life sciences at Amazon Web Services in Seattle, agrees. "Storing a data set on Amazon doesn't automatically enhance it," he says. But if scientists handle processing, curation and annotation, Amazon can be a secure, reliable data host, he says. It already houses several widely used

resources, including raw data from the Allen Brain Observatory and NeuroData.

Researchers can typically upload their data to life-science image databases at no cost, because storage, curation and maintenance are often funded by grants or other benefactors. Shuichi Onami, a developmental biologist at the RIKEN Center for Biosystems Dynamics Research in Kobe who founded the SSBD database, obtained funding from institutions including RIKEN; the Japan Science and Technology Agency; and the nation's Ministry

"We do all of the boring infrastructure to make sure those data persist."

of Education, Culture, Sports, Science and Technology. The database is "completely free" to the user, says Onami. Now he is expanding it beyond developmental biology, to include any biological data set that contains spatiotemporal information, as well as static images taken with state-of-the-art technologies.

It's also generally free to download data sets, and often to reuse and republish them: repositories frequently use Creative Commons licences that make availability transparent.

Pick and choose

Databases differ in the sizes of files they will accept, whether images must be linked to a published study, and their research focus. If your scientific community already has a specialized data house, Vogelstein recommends using that.

But there are general repositories. Figshare, for instance, accepts any kind of data, up to 5 gigabytes per file, for free. It can sometimes raise the limit, says founder Mark Hahnel – the biggest Figshare data set measures in terabytes. (Figshare is owned by Digital Science, a firm operated by the Holtzbrinck Publishing Group, which has a share in *Nature's* publisher, Springer Nature.) Other free, catch-all type services include Zenodo and Dryad.

Figshare also has contracts with universities, funders and publishers (including Springer Nature), which pay an annual fee for extra benefits. Last year, it set up a repository of data from research funded by the US National Institutes of Health (NIH), which now expects its grant recipients to make their results freely available. The site is meant for data that don't fit neatly into subject-specific banks, and currently hosts dozens of data sets. Unlike with the standard Figshare service, the NIH has control over the repository: what kinds of content are allowed, for example, and what kinds of metadata are required. NIH grantees benefit from assistance with metadata for their submissions, among other features.

The IDR databases – there is one for images

of cells and one for tissues – are tightly curated, says Swedlow. He and the other curators seek reference data sets linked to publications, such as results from large screening studies that would be of use to a wide audience. They ensure that the data are properly formatted and annotated with relevant metadata, such as information on the microscope used and experimental treatments applied.

Last July, EMBL-EBI announced a service called the BioImage Archive, which will host both the IDR and EMPIAR, as well as the more general BioStudies database. The institute will support further curated, community-specific databases in future, says Jo McEntyre, associate director for services at EMBL-EBI. With support from EMBL and the funding agency UK Research and Innovation, the BioImage Archive will be maintained for "as long as it's scientifically useful", she promises. Figshare, Hahnel says, "will persist forever" – although he admits the contract guarantees only a decade. "We do all of the boring infrastructure to make sure those data persist," he says.

Other people's data

These services make it easier to find, share and store big data sets. But as with DNA and protein databases, the hope is that image-surfers will find new science in others' data.

Demonstrating this potential, Swedlow and his colleagues combed images from three separate studies of cell elongation in the IDR. Two were from the human cancer-cell line HeLa; one was in fission yeast; all three imaged cells missing a variety of genes. "Each study gets different results, but they're related," says Swedlow. Together, these studies allowed him and his team to identify a larger, more complete network of genes involved in elongation than they could get from any one data set alone⁴.

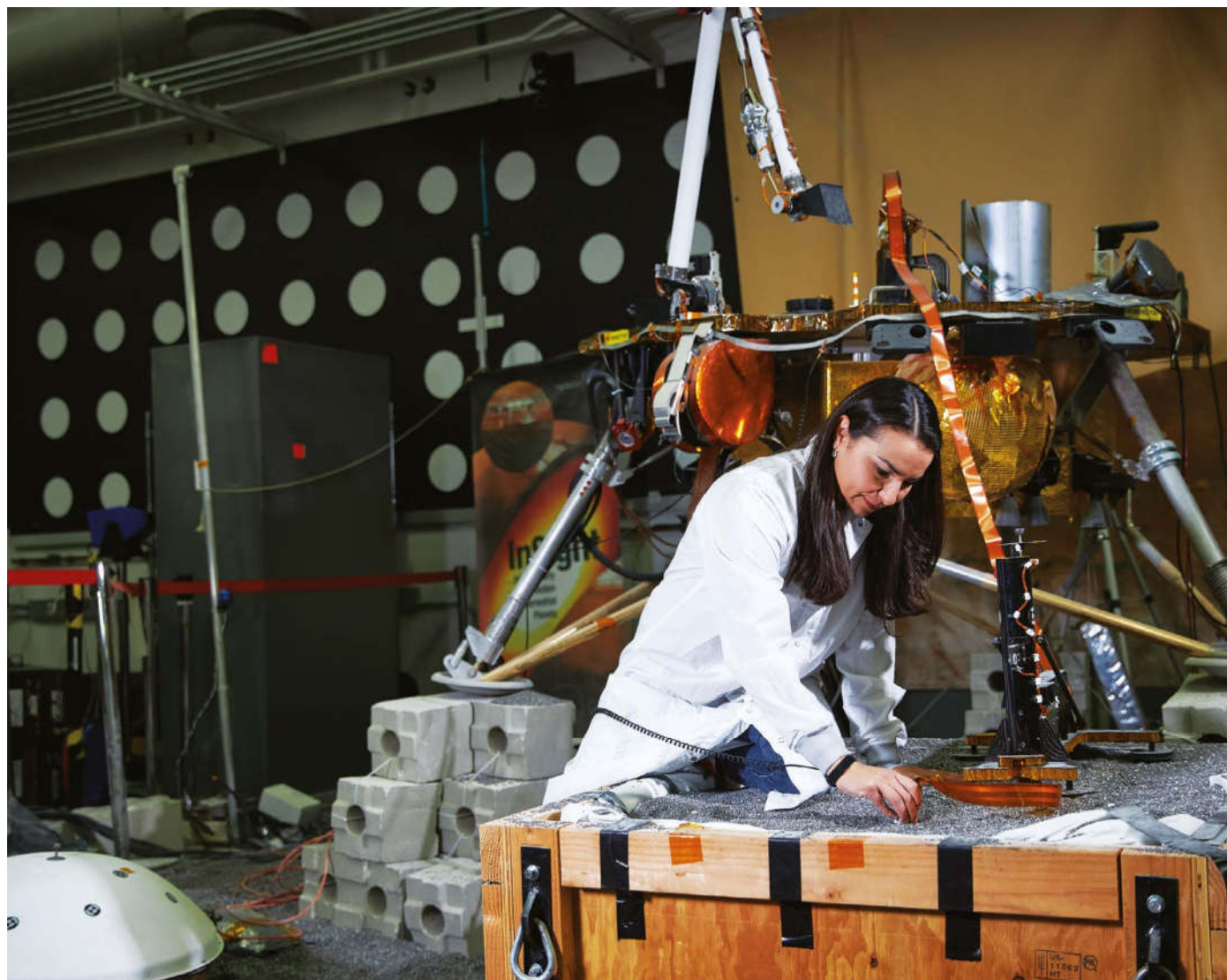
A study⁵ posted on the arXiv preprint server last year reports that of almost 532,000 journal articles published by PLOS and BioMed Central, those that linked to a data repository had up to a 25% higher citation impact than those that didn't.

With time to mature, image databases could yield more than just one-off discoveries, Swedlow says. After all, bioinformatics itself grew out of DNA archives.

"Hopefully," he says, "we end up stimulating the development of whole fields."

Amber Dance is a freelance science journalist near Los Angeles, California.

1. Zivanov, J., Nakane, T. & Scheres, S. H. W. *IUCr* <https://doi.org/10.1107/S2052252520000081> (2020).
2. Dorkenwald, S. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2019.12.29.890319> (2019).
3. McDole, K. et al. *Cell* **175**, 859–876 (2018).
4. Williams, E. et al. *Nature Meth.* **14**, 775–781 (2017).
5. Colavizza, G., Hrynaskiewicz, I., Staden, I., Whitaker, K. & McGillivray, B. Preprint at <https://arxiv.org/abs/1907.02565> (2019).



Where I work Marleen Martinez Sundgaard

Photographed for *Nature* by
Rocco Ceselin.

As a NASA engineer, I have the coolest job: I play with robots in a 'sandbox'. Anything a lander or rover on Mars does, we first test here on Earth in our sandbox filled with crushed garnet. We need a dust-free space, and garnet is so hard that it creates no dust. We use bright lights to mimic the sunlight on Mars, so pictures taken here are comparable to those captured on the red planet.

For the past few years, I've been working on the international InSight mission, which is studying the interior structure of Mars. The planet doesn't seem to have tectonic plates, so the surface that the landers are digging into has the same composition that it had one million years ago.

Before the mission launched on 5 May 2018, we spent about one year testing a full-scale model, shown behind me. We needed to ensure that the lander would be able to use all its equipment no matter what angle it landed at relative to the surface. We tested all possible angles. On 26 November 2018, InSight landed at a perfect two-degree tilt. We needed to troubleshoot how InSight

deploys its equipment, so we set up the sandbox like the landing site using a process called 'marsforming'. We used augmented-reality headsets, loaded with a digital map based on pictures taken by the lander. If you look down with one on, you are standing on 'Mars'. That brought tears to my eyes.

The lander's 'mole', a heat probe that is supposed to hammer five metres into the surface, is malfunctioning, so we needed to deepen the sandbox to find a fix. We raised the lander model on blocks and added an extra crate of garnet to make a small, deep space in which I could position a backup heat probe, as I'm shown doing, to test solutions to get the mole on Mars underground.

As a five-year-old, I wanted to become the first astronaut to walk on Mars. Since 2008, I've been applying for the astronaut corps. For now, I have the next-best thing.

Marleen Martinez Sundgaard is the lead systems test-bed engineer for the InSight and Psyche missions at the NASA Jet Propulsion Laboratory in Pasadena, California. **Interview by Amber Dance.**

Life and death decisions of autonomous vehicles

<https://doi.org/10.1038/s41586-020-1987-4>

Received: 2 May 2019

Accepted: 26 September 2019

Published online: 4 March 2020

 Check for updates
Yochanan E. Bigman^{1✉} & Kurt Gray¹Arising from: E. Awad et al. *Nature* <https://doi.org/10.1038/s41586-018-0637-6> (2018)

How should self-driving cars make decisions when human lives hang in the balance? The Moral Machine experiment¹ (MME) suggests that people want autonomous vehicles (AVs) to treat different human lives unequally, preferentially killing some people (for example, men, the old and the poor) over others (for example, women, the young and the rich). Our results challenge this idea, revealing that this apparent preference for inequality is driven by the specific ‘trolley-type’ paradigm used by the MME. Multiple studies with a revised paradigm reveal that people overwhelmingly want autonomous vehicles to treat different human lives equally in life and death situations, ignoring gender, age and status—a preference consistent with a general desire for equality^{2–4}.

The large-scale adoption of autonomous vehicles raises ethical challenges because autonomous vehicles may sometimes have to decide between killing one person or another^{5,6}. The MME seeks to reveal people’s preferences in these situations and many of these revealed preferences, such as ‘save more people over fewer’ and ‘kill by inaction over action’ are consistent with preferences documented in previous research^{7,8}.

However, the MME also concludes that people want autonomous vehicles to make decisions about who to kill on the basis of personal features, including physical fitness, age, status and gender (for example, saving women and killing men). This conclusion contradicts well-documented ethical preferences for equal treatment across demographic features and identities, a preference enshrined in the US Constitution, the United Nations Universal Declaration of Human Rights and in the Ethical Guideline 9 of the German Ethics Code for Automated and Connected Driving⁹.

We suggest that the MME finds preferences for inequality across lives because its methodology is relatively insensitive to preferences for equality. The MME uses trolley-type dilemmas that force people to choose between killing one person (or set of people) versus killing another person (or set of people). Because this paradigm assumes inequality (for example, should we program AVs to kill men or women?), it has difficulties revealing whether people prefer equality (for example, should we program AVs to ignore gender?).

What would happen if people indicated their ethical preferences in a revised paradigm, one that allowed AVs to treat different humans equally? We explored this possibility in study 1, in which people were randomly assigned to either a ‘forced inequality’ or an ‘equality allowed’ condition. Participants were drawn from two quasi-representative samples across two Western countries (US, $N = 1,174$; UK, $N = 1,178$).

The forced inequality condition was a simplified replication of the MME, testing whether participants thought autonomous vehicles should (1) kill group A (for example, elderly people) to save group B (for example, children) or (2) kill group B to save group A. As in the MME, we examined both personal features (for example, kill men versus

women) and structural features (for example, kill many people versus few people) in driving situations. However, unlike the MME—which used composite groups that simultaneously varied both personal and structural features—we examined each of these features individually (see Supplementary Information and https://osf.io/wy8tq/?view_only=e5907f552f5e4a8a901cbdd2d4c035f6 for details and data).

As Fig. 1 shows, results from the forced inequality condition closely match the global effects of the MME. Beyond the general value of replication¹⁰, this validates our paradigm: although we used a different sample and a simpler method, we obtained the same results as the MME.

The equality allowed condition was similar to the forced inequality condition, but with the addition of a third option, (3) treat the lives of groups A and B equally (for example, treat the lives of children and elderly people equally). As Fig. 1 shows, people overwhelmingly selected this option when it was available, revealing that they want autonomous vehicles to treat people equally. For example, when forced to choose between men and women, 87.7% chose to save women, but 97.9% of people actually preferred to treat both groups equally. See Supplementary Table 1 for full results.

Admittedly, it may be difficult to program a deep sense of egalitarianism into machines, but autonomous vehicles can functionally value human lives equally by simply ignoring (or failing to detect) features such as gender, age and social class. Restricting the ethical choice set of autonomous vehicles is consistent with emerging research revealing that people prefer autonomous machines not to make important ethical decisions^{11,12}. Ignoring personal features is also more consistent with the current technical capacities of AVs.

One question about our data is whether participants prefer the ‘treat equally’ option simply because it fails to mention killing. Study 2 ruled out this concern by replicating the equality allowed condition ($N = 843$ US participants from an online panel) with a modified third option: that autonomous vehicles should decide who to save and who to kill without considering their personal features. Consistent with study 1, people expressed a robust preference for AVs to treat people equally by ignoring personal features. For example, people preferred self-driving cars to not consider gender (92.6%), fitness (88.8%) or status (84.7%). The only substantial departure from study 1 was lawfulness: 53.1% of people preferred to spare law abiders over law breakers. See Supplementary Table 2 for full results.

Of course, AVs might sometimes have to choose between killing different sets of people, but these decisions can rely solely on structural rather than personal features. In study 3, participants ($N = 993$ US participants from an online panel) chose which of two autonomous vehicles should be allowed on the road: one that makes ethical decisions on the basis of the structural features revealed by the MME (for example,

¹Department of Psychology and Neuroscience, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ✉e-mail: ybigman@gmail.com

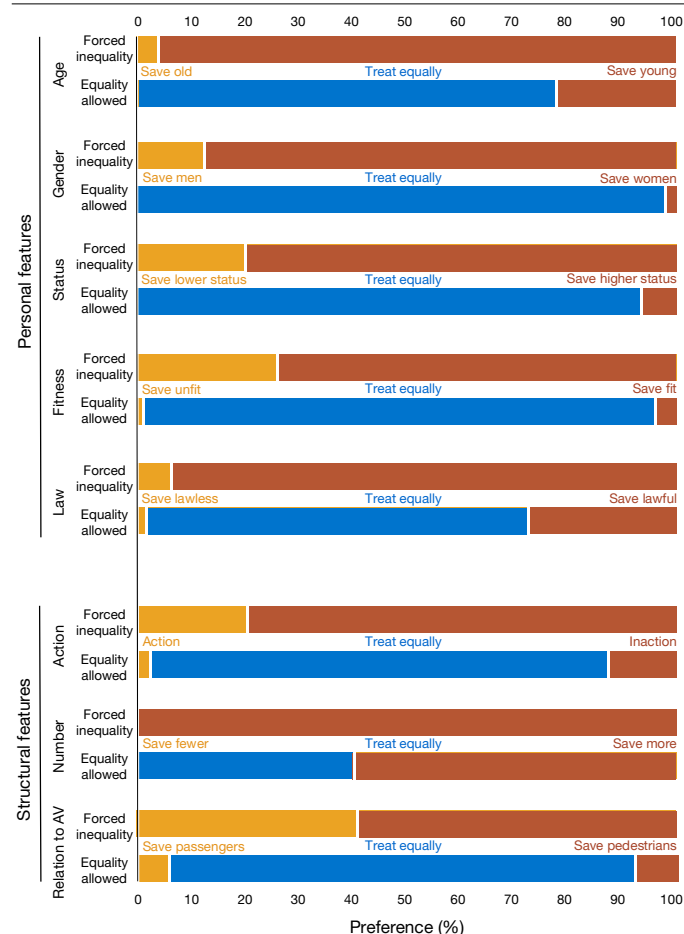


Fig. 1 | People's choices for how autonomous vehicles should be programmed to act in situations where human lives are at stake (study 1). Personal features reflect individual identity characteristics (for example, age and status) and structural features reflect characteristics of the situation. The forced inequality condition ($n=1,129$) replicates the MME, which makes people choose between two options, whereas the equality allowed condition ($n=1,223$) provides a third option of equal treatment. See Supplementary Fig. 1 for confidence intervals.

saving more people versus fewer, killing by inaction versus action), and another on the basis of both structural and personal features (for example, saving people based on age, gender, and status). Consistent with our predictions, 89.9% of participants chose the structural-features-only car, once again expressing a desire for AVs that ignore personal features in ethical dilemmas.

We note a number of caveats to our studies. Our samples were smaller than the millions who completed the MME. However, using quasi-representative samples in our main study (rather than a convenience sample) helps generalize the results to the populations of two large Western countries. We acknowledge that ethical preferences may vary across cultures, but our key point is that the current MME paradigm is relatively insensitive to preferences for equality, regardless of participant culture. Finally, we recognize that people often do discriminate on the basis of personal features, as sexism, classism, racism and ageism all illustrate. However, even people who implicitly act to perpetuate inequality often explicitly espouse ideas of equality¹³.

To frame the MME in a broader context, consider a thought experiment about some personal features not assessed by the MME—religion,

race, and disability. What might happen if the MME forced people to choose between black and white people? Aggregating people's decisions could reveal a racial bias¹³, but this would not mean that people want to share the road with racist autonomous vehicles. The same logic applies to the features that were included in the MME. Do people truly want to live in a world with sexist, ageist and classist self-driving cars? This thought experiment further suggests that aggregating across forced-choice preferences may not accurately reveal how people want autonomous vehicles to be programmed to act when human lives are at stake.

Although we must be careful about interpreting the results of the MME, we emphasize its value. Every methodology has limitations, and the MME reveals both basic moral cognitive processes and global preferences for saving lives in a forced-choice paradigm. More broadly, the MME highlights the important ethical questions posed by AVs—questions that society will soon need to address.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All materials, data and code used in the studies are available at https://osf.io/wy8tq/?view_only=e5907f552f5e4a8a901cbdd2d4c035f6.

- Awad, E. et al. The Moral Machine experiment. *Nature* **563**, 59–64 (2018).
- Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. Egalitarian motives in humans. *Nature* **446**, 794–796 (2007).
- Fehr, E., Bernhard, H. & Rockenbach, B. Egalitarianism in young children. *Nature* **454**, 1079–1083 (2008).
- Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
- Bonnefon, J.-F., Shariff, A. & Rahwan, I. The social dilemma of autonomous vehicles. *Science* **352**, 1573–1576 (2016).
- Li, J., Zhao, X., Cho, M.-J., Ju, W. & Malle, B. F. From trolley to autonomous vehicle: perceptions of responsibility and moral norms in traffic accidents with self-driving cars. *SAE Technical Paper* <https://doi.org/10.4271/2016-01-0164> (2016).
- Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R. & Hütter, M. Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *J. Pers. Soc. Psychol.* **113**, 343–376 (2017).
- Spranca, M., Minsk, E. & Baron, J. Omission and commission in judgment and choice. *J. Exp. Soc. Psychol.* **27**, 76–105 (1991).
- Luetge, C. The German Ethics Code for Automated and Connected Driving. *Philos. Technol.* **30**, 547–558 (2017).
- Gertler, P., Galiani, S. & Romero, M. How to make replication the norm. *Nature* **554**, 417–419 (2018).
- Bigman, Y. E. & Gray, K. People are averse to machines making moral decisions. *Cognition* **181**, 21–34 (2018).
- Bigman, Y. E., Waytz, A., Alterovitz, R. & Gray, K. Holding Robots Responsible: The Elements of Machine Morality. *Trends Cogn. Sci.* **23**, 365–368 (2019).
- Banaji, M. R. & Greenwald, A. G. *Blindspot: Hidden Biases of Good People* (Bantam Books, 2016).

Acknowledgements Y.E.B. acknowledges support from the National Science Foundation (SMA-1714298). K.G. acknowledges support from the Charles Koch Foundation and from the National Science Foundation (BCS-1823944).

Author contributions Y.E.B. and K.G. planned the research, designed the experiment, collected the data, analysed the data and wrote the paper.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1987-4>.

Correspondence and requests for materials should be addressed to Y.E.B.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Qualtrics XM

Data analysis IBM SPSS Statistics Version 20

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data availability statement Full materials, data and code are available at https://osf.io/wy8tq/?view_only=e5907f552f5e4a8a901cbdd2d4c035f6

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	All three reported studies were quantitative.
Research sample	<p>Study 1: Three thousand and three people were recruited by Prolific in two nationally representative samples (on age, gender and ethnicity) one from the UK and one from the USA. After a few days of data collection the age criteria was loosened, such that older ages are still a little under-represented (e.g., for people older than 58 years old, 328 instead of 467 in the US sample and 446 instead of 463 in the UK sample).</p> <p>In the UK representative sample (N = 1503), 772 were female and 731 male, 271 participants were between the ages of 18 and 27, 263 between 28 and 37, 282 between 38 and 47, 240 between 48 and 57, and 446 participants older than 58. One hundred and fifteen participants were Asian, 55 black, 31 mixed, 24 other and 1278 were white. Four of the responses were empty, such that the final sample size was 1499.</p> <p>In the US representative sample (N = 1500) 769 were female and 731 male, 339 participants were between the ages of 18 and 27, 327 between 28 and 37, 258 between 38 and 47, 248 between 48 and 57, and 328 participants older than 58. 96 were Asian, 197 black, 37 mixed, 30 other and 1140 were white.</p> <p>Study 2: One thousand and four people were recruited via Amazon's Mechanical Turk (429 male, 566 female, 9 other/preferred not to disclose; Age: M = 35.00, SD = 12.22).</p> <p>Study 3: One thousand and nine people were recruited via Amazon's Mechanical Turk (433 male, 570 female, 6 other/preferred not to disclose; Age: M = 37.26, SD = 12.80).</p>
Sampling strategy	<p>Study 1 used stratified sampling. Studies 2-3 used convenience samples.</p> <p>Sample size: We wanted to far exceed typical power recommendations, and given that isolating the true proportion of the population is important, believed 3000 participants would keep the standard error of the mean sufficiently low for our main study (Study 2), and 1000 for the additional Studies (Studies 2 and 3).</p>
Data collection	Data was collected on Qualtrics XM through online panels such as Prolific (Study 1) and Amazon's Mechanical Turk (Studies 2-3).
Timing	Study 1: The UK sample was collected between April 18th and April 23rd 2019. The US sample was collected between April 24th and April 30th. Data for Study 2 was collected on July 18th 2019. Data for Study 3 was collected on April 15th 2019.
Data exclusions	<p>All exclusions were per-registered (links to the per-registration appear in the Supplemental information).</p> <p>Study 1: Participants completed three attention checks. In the first attention check they were asked what day was yesterday and what they asked for breakfast. In the second attention check participants were shown three sliders, marked X, Y and Z. They were asked to set X on 15, Y to be greater than X and evenly divisible by 10, and Z to be larger than Y. In the third attention check participants were asked if they answered questions about how a self-driving car or a human driver, and if they had an option of having people treated equally. Six hundred and forty seven participants failed at least one of the attention checks and were excluded from the analysis as planned in the pre-registration.</p> <p>Study 2: Participants completed two attention checks. In the first attention check they were asked what day was yesterday and what they asked for breakfast. In the second attention check participants were asked if they answered questions about how a self-driving car or a human driver, and if they had an option of having people treated equally. One hundred and fifty seven participants failed at least one of the attention checks and were excluded from the analysis as planned in the pre-registration.</p> <p>Study 3: Participants were asked what day was yesterday and what they asked for breakfast. Sixteen participants failed this attention check and were excluded from the analysis.</p>
Non-participation	NA
Randomization	Randomization was done with the "randomize" function in Qualtrics.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

See above

Recruitment

Participants were recruited via online panels. While these panels perhaps do not reflect the general population perfectly, there is ample evidence that they provide good and valid results:
 Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48(1), 400–407. <https://doi.org/10.3758/s13428-015-0578-z>
 Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, 23(3), 184–188. <https://doi.org/10.1177/0963721414531598>
 Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, 26(3), 213–224. <https://doi.org/10.1002/bdm.1753>

Ethics oversight

The University of North Carolina at Chapel Hill IRB

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Reply to: Life and death decisions of autonomous vehicles

<https://doi.org/10.1038/s41586-020-1988-3>

Published online: 4 March 2020

Edmond Awad^{1,2}, Sohan Dsouza¹, Richard Kim¹, Jonathan Schulz³, Joseph Henrich⁴, Azim Shariff^{5*}, Jean-François Bonnefon^{6*} & Iyad Rahwan^{1,7,8*}

Replying to: Y. E. Bigman & K. Gray. <https://doi.org/10.1038/s41586-020-1987-4> (2020)

In ‘The Moral Machine experiment’ (MME)¹, we argued that policymakers would benefit from being aware of citizens’ preferences regarding the behaviour of autonomous vehicles in critical situations—situations in which an autonomous vehicle cannot save everyone, but can still decide to save one group of road users or another. In the accompanying Comment², Bigman and Gray make the important point that the way we measure these preferences can affect the results we obtain.

Actual consumer choices cannot yet be recorded. If we want the ethics of these vehicles to be decided before they hit the market, we can only collect stated preferences, based on hypothetical choices. The MME used a standard method for collecting stated preferences between multidimensional outcomes: Users chose between pairs of unavoidable accidents—which varied along multiple dimensions—and the importance of each dimension was statistically extracted from their choices using conjoint analysis³. Typical surveys can only do this for a few dimensions, because of the exponential increase in required sample size for every additional dimension. Given the unusual scale of the MME, we were able to investigate nine dimensions simultaneously.

Bigman and Gray adopted a different method. Rather than having users go through multiple pairs of nine-dimensional outcomes, they asked eight separate questions about general policy preferences, one per dimension (the human–nonhuman dimension was not used in their survey). For example, they asked: should self-driving cars be programmed to (1) kill children and save elderly people, (2) kill elderly people and save children, or (3) treat the lives of children and elderly people equally?

Bigman and Gray report that for all but one question—saving many versus few—the most frequent response was (3). For example, about 80% of participants said that self-driving cars should ‘treat the lives of children and elderly people equally’.

These results roughly agree with the Moral Machine results on some dimensions (for example, the weak preference for inaction), and disagree on others (for example, the preference for saving children), but the differences between the two methods, measures and statistical analyses make any direct comparison difficult. The two different methods may differently tap a single, stable set of preferences or they may elicit from respondents different facets of fragmented, inconsistent preferences that have yet to be solidified. Each approach comes with its own limitations, and its own usefulness. The Moral Machine approach allows us to measure the weight of different moral priorities when pitted against each

other, rather than considered in isolation; but participants cannot explicitly state that one dimension (for example, age) should not be taken into account. Of course, since each scenario involved at least two moral dimensions, respondents could avoid making decisions based on dimensions they felt should not be programmed into the cars. Participants who believed that the vehicle should be blind to age, for instance, could endeavour to be systematically blind to age themselves in how they responded to the scenario pairs. Had millions of participants made this choice, this would have statistically resulted in an absence of a preference for age, and it would have ranked at the bottom of the list of the nine moral dimensions we tested. It remains, however, that individuals had no opportunity to explicitly express this preference for equality.

The approach used by Bigman and Gray does offer participants the opportunity to explicitly express a preference for equality. One limitation of this approach is that measurement becomes sensitive to social desirability, experimental demands and framing effects (which is not to say that other methods do not have this problem). For example, consider the phrasing of the three response options above, and note how the word ‘kill’ disappears from the third option, making it instantly more attractive at a surface level. The first two options clearly describe trade-offs, whereas the third option only has positive connotations. We could suggest an opposite framing for the third option: ‘the self-driving car should indiscriminately kill children and elderly people’. This is as valid a description as the one used by Bigman and Gray, but it seems less attractive in this negative framing. Indeed, in their study 2, Bigman and Gray used a framing that stands somewhere in between the positive framing used in study 1 and the negative framing we suggest above, and this intermediate framing appeared to have an effect on the results: for half of the questions, the frequency of the ‘equality’ response decreased by 16 percentage points to 27% (as can be seen by comparing their Supplementary Table 1 and Supplementary Table 2).

We should note that an unpublished portion of the MME used a third method—one similar to that of Bigman and Gray, but one that avoided this loaded language confounder. After making 13 decisions, users had the option to ‘help us better understand (their) decisions’. Users who agreed were taken to a page where they could position one slider for each of the nine dimensions explored by the Moral Machine. For example, one slider showed a baby on the left side, an elderly person on the right side, and was labelled ‘Age preference’. Users could move the slider to express how important this dimension should be—more to the left if they wanted to save younger lives,

¹The Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Department of Economics, University of Exeter Business School, Exeter, UK. ³Department of Economics, George Mason University, Fairfax, VA, USA. ⁴Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. ⁵Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada. ⁶Toulouse School of Economics (TSE-M), CNRS, Université Toulouse Capitole, Toulouse, France. ⁷Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁸Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany. *e-mail: shariff@psych.ubc.ca; jean-francois.bonnefon@tse-fr.eu; rahwan@mpib-berlin.mpg.de

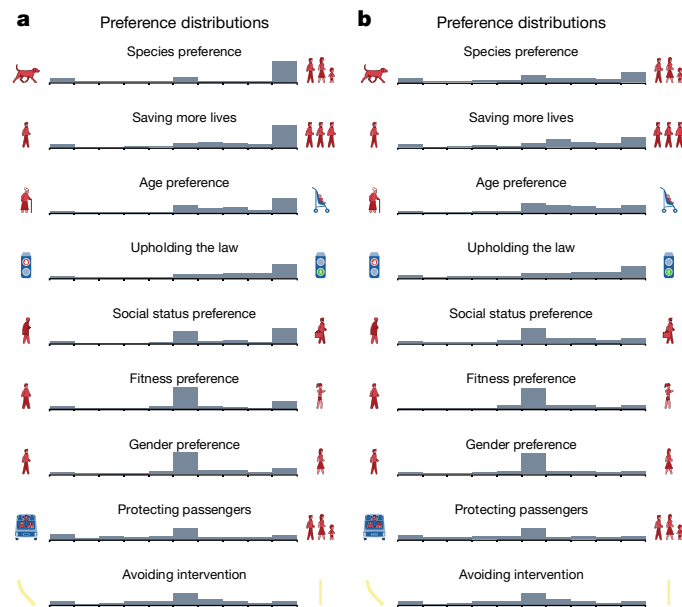


Fig. 1 | Distribution of explicit preferences stated by Moral Machine users. Sliders were presented with a default position determined by the responses users gave to the Moral Machine ‘judge’ mode. **a**, Preferences of users who moved at least one slider from its original position (585,531 users; >99% of the users). **b**, Preferences of users who changed sliders from their original position (range: 190,862–581,496 users). In both cases, only row 5 (social status preference) shows a clear gap between the preferences extracted from the Moral Machine¹ and the preferences explicitly expressed by users.

more to the right if they wanted to save older lives. Importantly, this method did give participants the option to treat the lives of children or elderly people (or men or women, or humans or pets) equally; participants could easily express such a preference by positioning the slider at the midpoint of the scale. This is, in essence, the method used by Bigman and Gray—except that it uses a continuous measure rather than a three-point scale and does not use a textual description for the midpoint of the scale.

The original position of the sliders was not systematically the middle point of the scale, but rather a rough estimation of the preference of each individual user based on their responses to the Moral Machine. Thus, users had the opportunity to move sliders if they disagreed with the estimation. More than 99% of users who saw the slider page moved at least one slider from its original position. Figure 1a shows the final position of all sliders for these 585,531 users, thus reflecting their choices when given the option of explicitly valuing all lives equally. Figure 1b shows the final position of each slider only for those users who actually moved it. This is a stronger test, since it restricts the data to the responses of users who actively expressed a preference.

Both figures tell a similar, three-part story. At the top of each figure, we can see that four preferences that were estimated as strong in the MME (saving humans, saving more lives, saving younger lives and saving pedestrians who cross legally; Fig. 2) are confirmed as strong. For these four dimensions, the distributions of responses are clearly skewed, and the modal response is not equality. At the bottom of each figure, four preferences that were identified as weak in the MME (inaction, saving pedestrians, saving fit characters and saving women) are confirmed as weak. The modal response for these dimensions is indeed equality.

Only for one dimension do we find a clear gap between the preferences extracted from the Moral Machine and the preferences explicitly expressed by users. Whereas users’ scenario-based choices indicated a

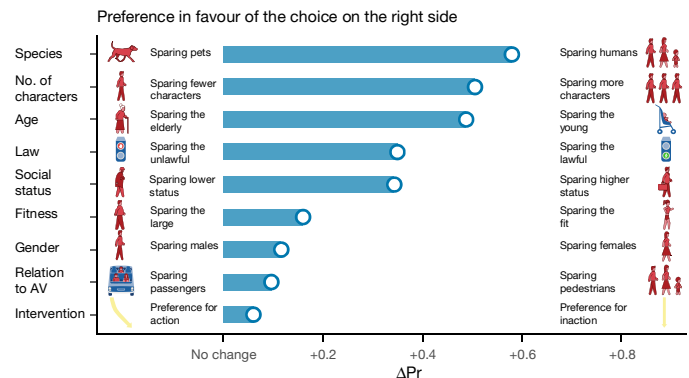


Fig. 2 | Preferences extracted from the conjoint analysis of the Moral Machine dataset. This figure is a simplified version of Fig. 2a from the MME¹. The x-axis shows the average marginal causal effect for each preference. In each row, ΔPr is the difference between the probability of sparing characters possessing the attribute on the right, and the probability of sparing characters possessing the attribute on the left, aggregated over all other attributes ($n = 35.2 \times 10^6$).

preference for saving high-status characters over low-status characters, their expressed preference on the sliders was to treat them equally. Here we see the value of giving people the opportunity to express an explicit preference: While their scenario-based choices may well show an implicit bias against lower-status victims, the users would probably be unhappy if this bias was actually acted on. Of course, it is extremely unlikely that policymakers would propose that autonomous vehicles should discriminate on the basis of social status, but we can still remain vigilant for other gaps between implicit biases and explicit preferences for equality, whenever they concern characteristics that may enter policy debates.

Self-driving car fatalities are an inevitability, but the type of fatalities that ethically offend the public and derail the industry are not. As a result, it seems important to anticipate, as accurately as we can, how the public will actually feel about the ethical decisions we program into these vehicles. Since any method used to collect these preferences will have its own biases and limitations, the methodological diversity advocated by Bigman and Gray, and the broad involvement of psychologists more generally, will be critical to reaching that goal.

Methods

Ethical compliance

This study was approved by the Institute Review Board at Massachusetts Institute of Technology. The authors complied with all relevant ethical considerations. Participants were briefed on the purpose of the study and were given the chance to opt out from having their data used.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data and code that can be used to reproduce Figs. 1 and 2 are available at <https://bit.ly/2VKyMhJ>.

1. Awad, E. et al. The Moral Machine experiment. *Nature* **563**, 59–64 (2018).
2. Bigman, Y. E. & Gray, K. Life and death decisions of autonomous vehicles. *Nature* <https://doi.org/10.1038/s41586-020-1987-4> (2020).
3. Hainmueller, J., Hopkins, D. J. & Yamamoto, T. Causal inference in conjoint analysis: understanding multidimensional choices via stated preference experiments. *Political Anal.* **22**, 1–30 (2014).

Acknowledgements J.-F.B. acknowledges support from the ANR-Labex Institute for Advanced Study in Toulouse, the ANR-3IA Artificial and Natural Intelligence Toulouse Institute, and the grant ANR-17-EURE-0010 Investissements d'Avenir. I.R. acknowledges funding from the Ethics & Governance of Artificial Intelligence Fund.

Author contributions I.R., A.S. and J.-F.B. planned the research. I.R., A.S., J.-F.B., E.A. and S.D. designed the experiment. E.A. and S.D. built the platform and collected the data. E.A., S.D., R.K., J.S. and A.S. analysed the data. E.A., S.D., R.K., J.S., J.H., A.S., J.-F.B. and I.R. interpreted the results and wrote the paper.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1988-3>.

Correspondence and requests for materials should be addressed to A.S., J.-F.B. or I.R.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data was collected through the Moral Machine website (www.moralmachine.mit.edu) which was built especially for the purpose of this study.

Data analysis

Data was preprocessed, analyzed, and visualized using Python (Jupyter 3.0), R (RStudio 3.4.1), and D3.js.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data and code that can be used to reproduce Figs 1 and 2 is available at the following link: <https://bit.ly/2VKyMhJ>

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study uses quantitative data collected from an online website designed as a randomized controlled multi-factorial design experiment. Part of the data is collected in a survey filled at the end.
Research sample	Research sample is Internet users who chose to visit the website and contribute to the data. Research sample has 2.3M users. The demographic information is only available for 585,565 users (aged between 15-75; 27% are females and 2% are "others").
Sampling strategy	The sample is self-selected. There was no power calculation. For Figure 2: sample size is 2.3M participants. For Figure 1: sample size for survey users is 585,565.
Data collection	Data was collected via the website http://moralmachine.mit.edu . Allocations of users to conditions was done automatically by the website, and so all researchers were blind to the experimental conditions. Data was stored in a MongoDB data base on a remote server.
Timing	Website was deployed on June 23rd, 2016. The data used in the analysis was collected continuously up until Dec 20th, 2017
Data exclusions	Similar to the MME paper, Figure 2 excluded responses for which the participant took more than 30 min. This resulted in the exclusion of 33,838 responses (out of 39.6M). Figure 1 excluded responses for which the participant did not change any of the nine sliders from their default values. This resulted in the exclusion of 34 responses (out of 585,565K).
Non-participation	Out of 2.86M completed sessions (13 scenarios), 43,979 sessions were opted out for. The number of participants dropping out (out of 2.3M) is hard to exactly know, but is comparable to the number of dropped out sessions. We did not collect the reasons to drop out.
Randomization	For Figure 1, randomization was done on the order of presentation of sliders. For Figure 2, similar to MME, users who visit the website get presented with 13 scenarios that are drawn from six main different conditions (2 scenarios from each condition + 1 fully random scenario). The six conditions vary the following aspect of characters: age, gender, fitness level, social status, number, and whether they are humans or pets. In conjunction with these six conditions, three main conditions were randomized: interventionism, relation to AV, and legality. Within each main condition, characters are sampled from a set of 20 characters (e.g. adult male, female athlete, homeless person, etc.).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	See above
Recruitment	Research sample is Internet users who chose to visit the website and contribute to the data. Thus, the sample is self-selected, and it is representative of a subset of the full population.
Ethics oversight	This study was approved by the Institute Review Board (IRB) at Massachusetts Institute of Technology (MIT).

Note that full information on the approval of the study protocol must also be provided in the manuscript.